PB82-108143

Performance of Trickling Filter Plants
Reliability, Stability, Variability

California Univ., Davis

Prepared for

Municipal Environmental Research Lab.
Cincinnati, OH

Sep 81

U.S. Department of Commerce
National Technical Information Service
NTIS

BEGIN FIRST LINE OF TEXT HERE
CENTER OF PAGE
TOP OF IMAGE AREA

EPA-600/2-81- 228
September 1981
PB82-108143

DROPPED HEAD BEGIN SECTIONS HERE

PERFORMANCE OF TRICKLING FILTER PLANTS

Reliability, Stability and Variability

by

Richard Haugh
Salar Niku
Edward D. Schroeder
George Tchobanoglous
University of California
Davis, California 95616

EPA Grant No. R805097-01

Project Officer
Jon Bender
Wastewater Research Division
Municipal Environmental Research
Laboratory
Cincinnati, Ohio 45268

MUNICIPAL ENVIRONMENTAL RESEARCH LABORATORY
OFFICE OF RESEARCH AND DEVELOPMENT
U.S. ENVIRONMENTAL PROTECTION AGENCY
CINCINNATI, OHIO 45268

REPRODUCED BY
NATIONAL TECHNICAL
INFORMATION SERVICE
U.S. DEPARTMENT OF COMMERCE
SPRINGFIELD, VA 22161

BEGIN LAST LINE OF TEXT
1/8"
BOTTOM OF IMAGE AREA OUTSIDE DIMENSION FOR TABLES AND ILLUSTRATIONS
EPA Form 2350-4 (4-80)
(PREVIOUSLY CIN. EPA FORM 383)
© U.S. GOVERNMENT PRINTING OFFICE 1980-660-630
TYPING GUIDE SHEET

TECHNICAL REPORT DATA

(Please read Instructions on the reverse before completing)

1. REPORT NO. EPA-600/2-81- 228
2. ORD Report
3. RECIPIENT'S ACCESSION NO. PB82 10814 3

4. TITLE AND SUBTITLE
Performance of Trickling Filter Plants: Reliability, Stability, Variability

5. REPORT DATE September 1981

6. PERFORMING ORGANIZATION CODE

7. AUTHOR(S)
Richard Haugh, Salar Niku, Edward D. Schroeder, and George Tchobanoglous

8. PERFORMING ORGANIZATION REPORT NO. N/A

9. PERFORMING ORGANIZATION NAME AND ADDRESS
Civil Engineering Department
University of California
Davis, California 95616

10. PROGRAM ELEMENT NO. P.E. 36B1C Task C1

11. CONTRACT/GRANT NO. Grant No. R805097-01

12. SPONSORING AGENCY NAME AND ADDRESS
Municipal Environmental Research Laboratory - Cinti, OH
Office of Research and Development
U. S. Environmental Protection Agency
Cincinnati, Ohio 45268

13. TYPE OF REPORT AND PERIOD COVERED Final 1979

14. SPONSORING AGENCY CODE EPA/600/14

15. SUPPLEMENTARY NOTES
Project Officer - Jon Bender (513) 684-7607

16. ABSTRACT

Effluent quality variability from trickling filters was examined in this study by statistically analyzing daily effluent BOD_5 and suspended solids data from 11 treatment plants. Summary statistics (mean, standard deviation, etc.) were examined to determine the general characteristics of those data. Distributions of most effluent data were skewed to the right, and daily suspended solids data generally exhibited more variation than daily BOD_5 data. Five probability distribution functions, chosen through experience and the literature, were tested to determine which would be best used to describe daily, 7-day average and 30-day average effluent data distributions. Three distributions, two parameter empirical, gamma and lognormal were found to be adequate with the lognormal being preferred because of ease of application. Daily effluent BOD_5 and suspended solids data were found to contain both random and non-random components. Weekly cycles were found in about half the plants studied and significant month to month variation in effluent quality was found in every plant. Effluent BOD and suspended solids concentrations were higher in winter than in summer when pooled data from all plants was examined. Multiple regression analysis was used to determine the effects of various process parameters on effluent quality. In general, primary effluent BOD_5 and suspended solids concentrations and wastewater temperature had the greatest effect. Variation due to measurement effort was estimated to be 5 to 70 percent and 11 to 78 percent for effluent BOD_5 and suspended solids values, respectively. Methods for incorporating statistical concepts into trickling filter design and operation were discussed.

17. KEY WORDS AND DOCUMENT ANALYSIS

a. DESCRIPTORS	b. IDENTIFIERS/OPEN ENDED TERMS	c. COSATI Field/Group
Sewage Treatment Trickling Filtration Effluents Reliability Stability Variability Statistical Analysis		13B

18. DISTRIBUTION STATEMENT Release to the Public	19. SECURITY CLASS *(This Report)* Unclassified	21.
	20. SECURITY CLASS *(This page)* Unclassified	22. PRICE

EPA Form 2220-1 (9-73)

DISCLAIMER

This report has been reviewed by the Municipal Environmental Research Laboratory, U.S. Environmental Protection Agency, and approved for publication. Approval does not signify that the contents necessarily reflect the views and policies of the U.S. Environmental Protection Agency, nor does mention of trade names or commercial products constitute endorsement or recommendation for use.

FOREWORD

The U.S. Environmental Protection Agency was created because of increasing public and government concern about the dangers of pollution to the health and welfare of the American people. Noxious air, foul water, and spoiled land are tragic testimonies to the deterioration of our natural environment. The complexity of that environment and the interplay of its components require a concentrated and integrated attack on the problem.

Research and development is that necessary first step in problem solution; it involves defining the problem, measuring its impact, and searching for solutions. The Municipal Environmental Research Laboratory develops new and improved technology and systems to prevent, treat, and manage wastewater and solid and hazardous waste pollutant discharges from municipal and community sources, to preserve and treat public drinking water supplies, and to minimize the adverse economic, social, health, and aesthetic effects of pollution. This publication is one of the products of that research and provides a most vital communications link between the researcher and the user community.

This report presents the results of a statistical analysis of the effluent quality of trickling filter plants. Normal variability is separated from variability due to operation and environmental factors. It is hoped that this information will be useful to engineers responsible for system design and water quality management.

Francis T. Mayo, Director
Municipal Environmental Research
Laboratory

BEGIN FIRST LINE OF TEXT HERE
CENTER OF PAGE
TOP OF IMAGE AREA
DROPPED HEAD: BEGIN SECTIONS HERE

ABSTRACT

This study examines effluent variability in trickling filters by statistically analyzing daily effluent BOD_5 and suspended solids data from 11 trickling filter wastewater treatment plants.

Summary statistics (mean, standard deviation, etc.) were examined to determine the general characteristics of those data. Distributions of most effluent data were skewed to the right, and daily suspended solids data generally exhibited more variation than daily BOD_5 data.

Two empirical probability distribution functions and the normal, log-normal, and gamma distributions were tested with the Kolmogorov-Smirnov goodness-of-fit test to determine which could best be used to describe daily, 7-day-average, and 30-day-average effluent data distributions. The log-normal, the two-parameter empirical, and the gamma distributions were found to be adequate for describing the data. The log-normal distribution was preferred because it was the easiest to use.

Daily effluent BOD_5 and suspended solids data were tested for randomness and were found to contain both random and non-random components. Weekly cycles and month-to-month fluctuations were tested for by using analysis of variance techniques. Weekly cycles were evident in data from about half the plants studied, and significant month-to-month changes in effluent quality were found in data from every plant. An annual cycle in BOD_5 and suspended solids data was evident when data from all 11 plants were pooled: Effluent concentrations were higher in the winter and lower in the summer.

Two causes of effluent variability were studied. First, multiple regression analysis was used to determine the effects of various process parameters on effluent quality. Results varied from plant to plant, but in general, changes in primary effluent BOD_5 and suspended solids concentrations and wastewater temperature appeared to have the greatest effect on effluent quality. Second, the effect of measurement error on data fluctuations was estimated. The variation caused by measurement error was found to be 5 to 78 percent and 11 to 78 percent of the total variation in effluent BOD_5 and suspended solids data, respectively.

Recommendations for incorporating stochastic concepts in trickling filter plant design and reduction of effluent data variation are discussed.

This report was submitted in fulfillment of Grant No. R805097-01 by The Department of Civil Engineering, University of California, Davis under the sponsorship of the U.S. Environmental Protection Agency. This report covers the period April, 1979 to September 1979, and work was completed as of September 1979.

BEGIN LAST LINE OF TEXT
5/8"
BOTTOM OF IMAGE AREA OUTSIDE DIMENSION FOR TABLES AND ILLUSTRATIONS
PAGE NUMBER

CONTENTS

FOREWORD . . . iii
ABSTRACT . . . iv
FIGURES . . . vii
TABLES . . . viii
ABBREVIATIONS . . . x
ACKNOWLEDGEMENTS . . . xii

1. **Introduction** . . . 1
    - Background . . . 1
    - Trickling Filter Processes . . . 5
    - Objectives . . . 8
2. **Conclusions** . . . 9
3. **Literature Review** . . . 11
    - Effluent Probability Distributions . . . 11
    - Randomness of Influent and Effluent Data . . . 12
    - Daily Cycles in Effluent Data . . . 13
    - Weekly and Annual Cycles in Effluent Data . . . 13
    - Studies Using Spectral and Harmonic Analysis to Detect Periodic Cycles . . . 14
    - Time Series and Deterministic Physical Models . . . 15
    - Sampling and Measurement Errors . . . 16
4. **Description of Data and Analytical Methods** . . . 17
    - Description of Data . . . 17
    - Treatment of Missing Data . . . 20
    - Analytical Methods . . . 20
5. **Results** . . . 40
    - Effluent Data Distributions . . . 40
    - Randomness of Effluent Data . . . 53
    - Annual and Weekly Patterns . . . 55
    - Effects of Process Parameters . . . 65
    - Variation Caused by Measurement Error . . . 66
6. **Discussion** . . . 68
    - Effluent Distributions . . . 68
    - Randomness of Effluent Data . . . 70
    - Annual and Weekly Patterns . . . 70
    - Effects of Process Variables on Effluent Concentrations . . . 72
    - Engineering Applications . . . 73
7. **Reliability and Stability of Trickling Filter Processes** . . . 76
    - Effluent Quality Distribution of Trickling Filters . . . 76
    - Reliability Concept - Background . . . 76
    - Verification of Results . . . 80
    - Stability of Trickling Filter Plants . . . 80

8. Recommendations 85
    Effluent Distributions 85
    Annual and Weekly Cycles in Effluent Data 85
    Effects of Process Variables on Effluent Concentration . . . 86
    Measurement Errors 86

References. 88

Appendices

A. Coefficients and statistical results for two-parameter models. . . . 91
B. Deviation of monthly averages from annual mean 96

FIGURES

Number **Page**

1 Two Alternatives for Reducing the Frequency of Exceeding a Specified Effluent BOD_5 Concentration 4

2 Trickling Filter Plant Schematics 6

3 Example of Kolmogorov-Smirov Goodness-of-Fit Test 26

4 Example of Counting Runs in a Data Series 28

5 Annual Pattern of Normalized Data Pooled from 11 Plants . . . 59

6 Weekly Pattern of Normalized Effluent Data Pooled from 9 Plants 60

7 Reliability versus Normalized Mean for Different Coefficient of Variation. 81

8 Variability of Effluent BOD as a Function of Standard Deviation. 83

9 Variability of Effluent Suspended Solids Concentration as a Function of Standard Deviation 84

TABLES

Number		Page
1	Trickling Filter Treatment Plants Studied	18
2	Daily Data Available From Each Plant	19
3	Statistical Analyses Performed and Corresponding Data Files	21
4	Reported Precision of BOD_5 and Suspended Solids Tests	35
5	Estimated Precision of BOD_5 and Suspended Solids Tests for Mean Concentrations between Published Values	36
6	Summary Statistics of Effluent BOD_5 and Suspended Solids Concentration	41
7	Coefficients and Statistical Results For One and Two Parameters Emperical Distribution Models, Daily Effluent BOD_5 Concentrations	45
8	Results of Kolmogorov-Smirnov Test for Effluent BOD_5 and Suspended Solids Concentrations	46
9	Average Maximum Percentile Differences	52
10	Results of Randomness Tests for Daily Effluent BOD_5 and Suspended Solids	53
11	Results of 2-Way Analysis of Variance For Effluent BOD_5 and Suspended Solids	56
12	Coefficients For Part I Regression Models	61
13	Results of Part II Regression Analysis	63
14	Variation Caused by Measurement Error	67
15	Sum of Part II Regression R^2 and Estimated Percentage of Variance Due to Measurement Errors	74

EPA Form 2350-4 (4-80)
(PREVIOUSLY CIN. EPA FORM 267)
TYPING GUIDE SHEET

16 COR as a Function of V_x and Reliability Based on True Log-Normal Distribution 78

17 Reliability as a Function of Normalized Mean and V_x of Effluent Concentrations 79

18 Comparison of Predicted and Measured Performance of Trickling Filter Process 82

EPA Form 2350-4 (4-70)
(PREVIOUSLY CIN. EPA FORM 287)

TYPING GUIDE SHEET

BEGIN FIRST LINE OF TEXT HERE
CENTER OF PAGE
TOP OF IMAGE AREA
DROPPED HEAD BEGIN SECTIONS HERE

SYMBOLS AND ABBREVIATIONS

a_1, a_2	regression coefficients
A_x	regression coefficient for annual mean concentration in one-and two-parameter empirical models
ATEMP	average daily ambient temperature
AVTEMP	average temperature = (ATEMP + WTEMP)/2
B_x	regression coefficient for annual standard deviation in two-parameter empirical models
BOD_5	5-day biochemical oxygen demand
BOD_i, BOD_p, BOD_e	concentration of BOD_5 in influent, primary effluent, and secondary effluent, respectively
C	estimated standard deviation of measurement error, expressed as a fraction of the mean value of the data
C	a limiting concentration of some effluent constituent specified in a discharge requirement
C_x	constant in one- and two-parameter empirical models
D	test statistic in Kolmogorov-Smirnov goodness-of-fit test
DTEMP	draft temperature = | ATEMP - WTEMP |
e	secondary effluent
EPA	U.S. Environmental Protection Agency
EPV	estimated percentage of variance resulting from measurement error
F	test statistic in F-test
g/m^3	grams per cubic meter
i	influent
mg/l	milligrams per liter
mgd	million gallons per day
m_x	arithmetic mean of sample
p	primary effluent
$p_i(C)$	probability of the occurrence of C
pH	negative $\log_{10}$ hydrogen ion concentration
Q	flow rate
Q_R	recycle flow rate
r	residual, the difference between the predicted and observed value of a variable
r	bivariate correlation coefficient
r^2	bivariate coefficient of determination
R	recycle ratio
R^2	multiple regression coefficient of determination
R7DB, R7DS	running 7-day average BOD_5 and SS, respectively
R30DB, R30DS	running 30-day average BOD_5 and SS, respectively
S_x^2	sample variance
S_x	sample standard deviation

BEGIN LAST LINE OF TEXT

PAGE NUMBER
EPA Form 2350-4 (4-80)
(PREVIOUSLY EIN. EPA FORM 287)
TYPING GUIDE SHEET

SS	suspended solids
SS_i, SS_p, SS_e	suspended solids concentration in influent, primary effluent and secondary effluent, respectively
T_{avg}	average of the ambient air and wastewater temperatures
T_d	draft temperature equal to the absolute value of the difference between the ambient air and wastewater temperatures
V_X	sample coefficient of variation
WTEMP	wastewater temperature
X_i	observed variable concentration
X_m	measured concentration of X in a sample
X_s	actual concentration of X in a sample
$\bar{X}$	sample arithmetic mean
X%ile	Xth percentile, $0 \leq X \leq 100$
Y	dependent variable in regression equation
$\hat{Y}$	predicted value of Y
Z	standardized normal variable
α_i	effect of ith month in analysis of variance model
β_j	effect of the jth day of the week in analysis of variance model
$\alpha\beta_{ij}$	interaction effect in analysis of variance model
ϵ	random error
ϵ_a	measurement error
μ	mean value of a variable
σ_X	standard deviation of X
σ_X^2	variance of X

EPA Form 2350-4 (4-80)
(PREVIOUSLY CIN. EPA FORM 267)
© U.S. GOVERNMENT PRINTING OFFICE 1980-080-670
TYPING GUIDE SHEET

BEGIN FIRST LINE OF TEXT HERE
CENTER OF PAGE
TOP OF IMAGE AREA
DROPPED HEAD: BEGIN SECTIONS HERE

ACKNOWLEDGMENTS

The authors gratefully acknowledge the assistance of the staffs of the various water quality agencies and wastewater treatment plants who generously provided data. Jon Bender, the project officer provided both suggestions and enthusiastic support throughout the project. His efforts are particularly appreciated. We also thank Marlene Pendergrass and Dinah Pfoutz for their careful typing.

This report, except for Section 7 was originally submitted by Richard S. Haugh as a Master's thesis in the Department of Civil Engineering, University of California, Davis.

BEGIN LAST LINE OF TEXT
3/8"
BOTTOM OF IMAGE AREA: OUTSIDE DIMENSION FOR TABLES AND ILLUSTRATIONS

PAGE NUMBER
EPA Form 2350-4 (4-80)
(PREVIOUSLY CIN. EPA FORM 287)
TYPING GUIDE SHEET

SECTION 1

INTRODUCTION

The quality and quantity of wastewater treatment plant effluents is never constant, but fluctuates hourly, daily, and seasonally. Variations in treatment plant performance have long been recognized, but until recently, most theoretical models and design procedures for wastewater treatment processes have been concerned with average performance only.

Because of Federal and State legislation mandating the improvement and maintenance of water quality in our nation's streams, rivers and lakes, there has been a continuing need in recent years to improve the performance of wastewater treatment processes. Development of deterministic models for activated sludge, trickling filters, and other treatment processes has been a major research effort over the past 20 years. These models, directed towards improving treatment plant performance, have been concerned mostly with reducing the average concentrations of effluent pollutants, as opposed to trying to reduce the variability of effluent concentrations.

As discharge requirements have become more stringent, it has become necessary to include concepts of variability into the theory, design, and regulation of wastewater treatment plants. For example, limitations on the average daily and weekly concentrations of pollutants have made it necessary not only to insure that average performance is within discharge limits, but also to reduce the amount of variation in performance.

A number of recent studies have examined effluent variability in wastewater treatment plants, but most of these studies have been concerned soley with activated sludge plant performance. (1,9,21,24,38,43,44) The purpose of this study is to investigate and characterize the nature and sources of variability of effluent quality in trickling filter wastewater treatment plants. Five-day biochemical oxygen demand (BOD_5) and suspended solids (SS) concentrations are the effluent quality parameters that are examined.

BACKGROUND

Most theoretical and laboratory work on wastewater treatment processes, and practically all studies of data from full scale treatment plants, have been based on deterministic, steady state concepts. Recently, however, dynamic and stochastic models have been developed for both the activated sludge and trickling filter processes. (30,35,37) Laboratory studies have been done on the activated sludge process under variable loading conditions, (18,23,26) and treatment plant influent and effluent data have been studied using sophisticated stochastic models. (7,9,44)

EPA Form 2350-4 (4-80)
(PREVIOUSLY CIN: EPA FORM 287)
TYPING GUIDE SHEET

The recent upsurge of interest in wastewater treatment plant variability is due partly to the fact that stricter discharge requirements have made improving process reliability a necessity. Also, Adams and Gemmel (2) have suggested that stochastic models of wastewater treatment processes and time series studies of treatment plant data did not appear until recently because the mathematical methods and extensive data required for such research previously were not available.

There exists a need to incorporate the concept of variability in virtually every area of the wastewater treatment field. Numerous authors have argued that stochastic concepts should be included in wastewater treatment plant planning, design, and regulation. (1,2,8,25) In the following sections, areas where consideration of treatment variability is important are discussed.

Discharge Requirements

Effluent variability is implicitly recognized in current federal secondary treatment standards plant discharge limits. The 7-day average maximums, which allows for the fact that short term fluctuations tend to be averaged out in the long run.

However, daily, weekly and monthly average concentrations limits are not explicitly stochastic because they set maximum values which cannot ever exceed (without causing a violation). These fixed limits, expecially daily maximums, are conceptually inconsistent with the stochastic nature of wastewater treatment processes.

Some authors have suggested that treatment plant discharge requirements be stated in explicitly probabilistic language. (2,7) An example of a probabilistic standard would be one in which the daily composite average BOD_5 concentration cannot exceed 70 g/m^3 more than two times in any consecutive 30 day period.

One difficulty with fixed daily maximums is that if the standard is set at a level which indicates some significant threshold of pollution, then the standard is almost inevitably violated from time to time. On the other hand, if the daily standard is set so that it is virtually never exceeded, then the limit must be so high that it is rather meaningless as an indicator of a threshold pollution level.

Discharge requirements should also be consistent such that one requirement does not make another requirement redundant. For example, if it were shown that the secondary treatment 7-day average BOD_5 effluent concentration limit is much more stringent than the 30-day average requirement, then the 30-day requirement would be redundant. It could be concluded that either the 7-day limit is too strict, the 30-day limit is too lenient or that the 30-day limit is unnecessary and should be dropped. Current EPA secondary treatment standards have been shown to be consistent for activated sludge plants. (21)

Effluent limitations are intended to protect receiving water quality. Because stream quality and flow rate are also variable, treatment variability, receiving stream variability, and the interactions between the two must be quantitively understood to maintain water quality in a cost effective manner[2]. The maximum sag caused by a treatment plant discharge to a stream is often smaller during the winter than in the summer because of higher flow rates, greater oxygen saturation and lower rates of biological activity. Thus, it may be possible in some cases to allow slightly greater pollutant discharges in the winter. However, other factors such as the sensitivity of biota to pollution and oxygen levels during the winter and seasonal patterns of stream use must also be taken into consideration when discharge requirements are set.

Design

Stochastic concepts are also useful in treatment plant design. Designers must consider the degree of variability to be expected, as well as the average performance, if discharge requirements are to be met. Currently, most design techniques are based on steady state assumptions and average constituents values are used for design parameters. One method of accommodating process variability in activated sludge plants is to oversize the treatment units. (16) However, the inexact nature of this approach can lead to unnecessary oversizing of facilities.

Another way to allow for process variability is to determine the expected effluent variation based on data from existing treatment plants. Porter and Boone (39) and Hovey et al (21) developed linear regression models to predict the distribution precentiles of effluent BOD and SS concentrations, based on the annual mean value of each parameter. Working backwards, designers can choose an effluent concentration not to be exceeded more than a prescribed percentage of the time and then calculate the required average value. Using these computed average values as design criteria, traditional design methods can then be used to size the treatment units. If the design averages are achieved, treatment variability should be within the expected limits.

The design approach discussed above results in a reduction of the occurrence of high effluent concentrations because the average concentration of pollutants is reduced. Thus, the entire distribution of effluent concentrations is shifted to a lower level, and fewer high values occur. In Figure 1a, the effect of lowering the average concentration of effluent values is illustrated. The two probability density curves in Figure 1 represent the distribution of effluent BOD concentrations for two different treatment plant designs. The area under each curve to the right of the discharge limit, C, is the frequency of occurrence of effluent concentrations greater than C. Both distributions have the same magnitude of variation, but alternative 2 has lower average concentration and a lower frequency of exceeding the discharge limit than alternative 1.

Possibly, a more cost effective approach to reducing the occurrence of high effluent concentrations is to reduce the variability of treatment processes, rather than improving the average performance of these processes. Reducing the variability has the effect of narrowing the distributions of effluent values. This is illustrated in Figure 1b. The distributions of effluent concentrations for both alternatives have the same average value, but the distribution for alternative number 3 has less variation. Consequently, the frequency of discharge violations for alternative 3 is less.

The use of equalization chambers and automatic process control have been proposed for the purpose of reducing effluent variability. (4,15,25,36,46) Equalization of influent flows reduces loading fluctuations; because the treatment process is operated under loading conditions that approach steady state conditions, fluctuation in process performance are reduced. Effluent flow equalization prior to discharge reduces effluent quality variation directly by mixing flows from different periods of process performance. Automatic control can be used to adjust process parameters to compensate for changes in loading conditions and process efficiency.

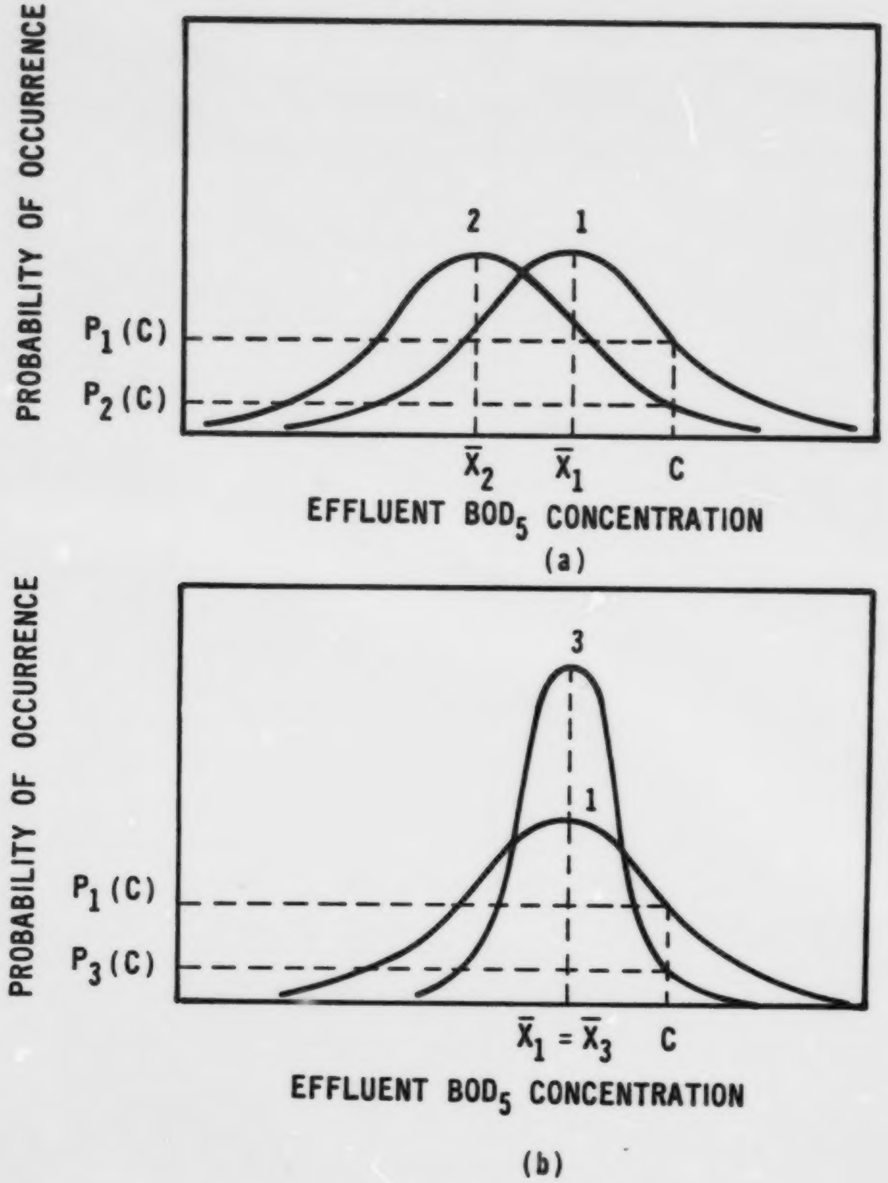


Figure 1. Two alternatives for reducing the frequency of exceeding a specified effluent BOD_5 concentration.

Regional Water Quality Planning

Considering the effects of effluent variability is also important in regional water quality planning. For example, Adams and Gemmel (1) have shown that one effect of centralizing all waste treatment at one regional plant is to cause greater fluctuations in stream quality as compared to treating wastes at several smaller plants. When a downward fluctuation in effluent quality occurs at a regional facility, the discharge for the entire wastewater system is of poor quality. However, the authors demonstrated that if several smaller facilities are used to treat the wastes from one city, the fluctuations in effluent quality at the different facilities tend to be uncorrelated. Thus, a downward fluctuation in effluent quality at one plant tends to be offset by an upward fluctuation in effluent quality at another, and the overall effect on stream quality tends to remain more constant.

The negative effects of treating wastewater at a single regional facility would be ameliorated if larger treatment plants tended to produce effluents of better average quality or produced effluents with smaller quality fluctuations. However, this does not appear to be the case. Hovey et al (21) have shown that larger activated sludge treatment plants actually tend to produce effluent concentrations distributions with slightly higher means and standard deviations than do smaller plants.

TRICKLING FILTER PROCESSES

The purpose of this study is to determine the variability in trickling filter effluent quality. To aquaint the reader with the trickling filter process, a brief description of the major components of trickling filter plants and a discussion of causes of effluent variability are presented below.

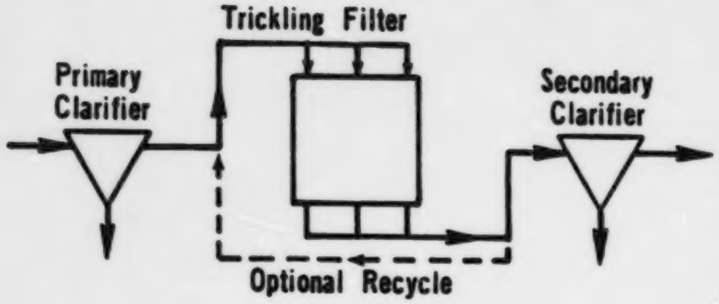
Process Description

Trickling filter systems are arranged in a variety of configurations. The major components of three typical process arrangements are illustrated in Figure 2.

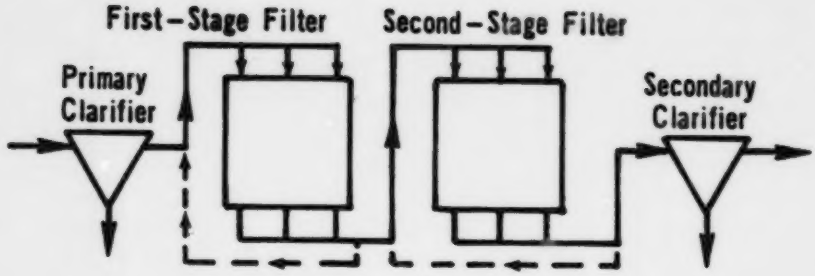
The two processes used in all trickling filter plant configurations are the trickling filters and clarifiers. Trickling filters, which remove organic matter and non-settleable colloidal solids from the waste stream, contain an inert (usually rock or plastic) medium that supports the growth of a biological slime. Wastewater is sprayed over the medium and trickles downward over the slime surface. Organic matter diffuses from the liquid into the slime layer and is oxidized by the microorganisms that make up the slime. Treated wastewater and portions of the slime layer that have sloughed from the medium are collected in an underdrain system, from which they are discharged to a clarifier for the removal of solids. The underdrain also serves as part of the ventilation system at the bottom of the filter.

Clarifiers are quiesent basins where settleable and floatable matter is removed. Primary clarifiers are used to reduce the solids and organic loading on the trickling filter units. Final and intermediate clarifiers remove biological solids which have sloughed from the trickling filters.

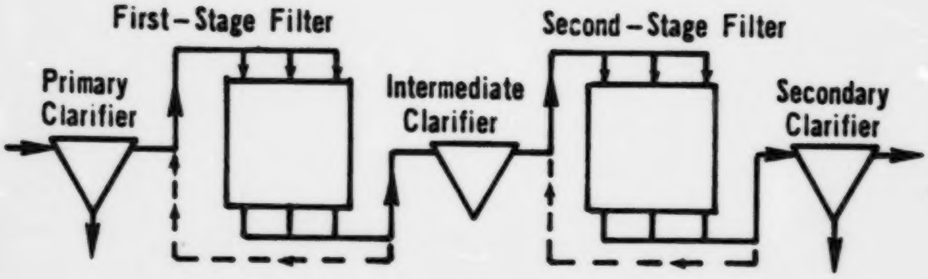
In many trickling filter plants, wastewater is recycled back through the filters (see Figure 2). Recirculation dilutes high-strength influent wastes, and if biological


Trickling Filter
Primary Clarifier
Secondary Clarifier
Optional Recycle

(a) One-Stage


First-Stage Filter
Second-Stage Filter
Primary Clarifier
Secondary Clarifier

(b) Two-Stage


First-Stage Filter
Second-Stage Filter
Primary Clarifier
Intermediate Clarifier
Secondary Clarifier

(c) Two-Stage With Intermediate Clarifier

Figure 2. Trickling filter plant schematics.

solids are recycled, a higher rate of liquid phase biological oxidation results. The higher flow rate through the filter which results from recirculation reduces ponding by controlling the buildup of slime in the medium. Also, the occurrence of flies and objectionable odor are reduced because the surface of the filter medium is continually washed.

Sources of Effluent Variability in Trickling Filter Plants

Effluent variability in trickling filter plants is caused by a number of environmental, operational, and loading factors. Factors which cause fluctuations in trickling filter and clarifier performance are discussed below.

Fluctuations in Trickling Filter Performance—The wastewater flow rate through the trickling filter controls the thickness and velocity of the liquid film on the biological slime. An increase in flow rate increases the film thickness which in turn inhibits oxygen transfer through the film. Higher liquid velocities reduce the contact time for diffusion of organics from the liquid into the slime. Thus, the overall effect of increased flowrate is a higher filter effluent BOD concentration for a given influent BOD concentration.

An increase in the concentration of organic matter entering the filter generally causes an increase in the filter effluent organic concentration. Increased organic concentration increases the mass loading rate on the filter and raises the concentration of organics in the liquid film. The increased concentration of organics causes food and nutrients to penetrate deeper into the slime layer, which increases the reaction rate per unit area unless the maximum reaction rate has been reached or some other factor, such as oxygen transfer, has become rate limiting. Although the mass removal rate increases, it does not increase as much as the increase in mass loading, and the net result is a higher filter effluent organic BOD concentration.

Air and wastewater temperature both effect treatment efficiency. A temperature increase in the liquid film and biological slime increases transport and reaction rates, thus improving treatment efficiency. In conventional filters, the supply of air (and thus oxygen) to the microorganisms is controlled by the relationship between air and wastewater temperature. The greater the difference between air and water temperature, the greater the draft through the filter.

Other factors that affect filter operation are wastewater pH and the presence of toxic substances. The optimal pH for microbial growth in conventional biological treatment is between 6.5 and 8.5. If the pH is considerably outside this range, treatment efficiency will decrease. (41) Toxic substances in the influent can also retard biological activity and reduce treatment efficiency.

Trickling filters are quasi-natural biological processes, and there is evidence that slime growth-decay cycles occur locally within the filter, independent of fluctuations in flow rate, organic concentration or other parameters. (22) It is not known whether these cycles cause significant variations in the overall performance of the filter.

Fluctuations in Clarifier Performance—The most important factors which affect clarifier performance are the concentration and composition of incoming suspended matter and the flow rate through the tank. An increase in flow rate decreases the time available for particles to settle to the bottom of the tank and thus increases the concentration of particles in the clarifier effluent.

Density, shape, and size all affect the settling velocity of particles. An increase in the settling velocity of the suspended particles entering the clarifier results in an increase in the percentage of particles removed.

Generally, higher concentrations of suspended matter in the clarifier influent result in higher effluent concentrations because there is an increase in the concentration of particles with settling velocities too low to be removed.

Floculation and straining of particles also affect suspended solids removal in clarifiers, but these mechanisms cannot be explained simply in terms of clarifier influent strength and flow rate.

Other factors can influence clarifier performance. Improperly operated sludge scrapers can cause disruptive currents in the tank and settled solids that remain on the bottom of the tank too long can become anaerobic and produce gas bubbles that float solids to the water surface.

OBJECTIVES

The purpose of this study is to investigate the characteristics and some possible sources of variability in final effluent quality in trickling filter wastewater treatment plants. Daily performance data have been obtained from 11 treatment plants and analyzed statistically to investigate the following questions:

1. How can the distributions of effluent BOD_5 and SS concentrations from trickling filter plants be characterized?
2. Are trickling filter effluent BOD_5 and SS concentrations data random?
3. If trickling filter plant effluent BOD_5 and SS data are not random, do the data follow any annual or weekly cycles?
4. What are the effects of influent, environmental, and operational parameters on effluent variability?
5. What is the effect of measurement error in effluent data variability?

It would be worthwhile to study the variability of each process. However, data from between the trickling filter and final clarifier are not generally available. Therefore, in this study only the variability of the final effluent is examined.

SECTION 2

CONCLUSIONS

Daily performance records from 11 trickling filter wastewater treatment plants have been statistically analyzed to determine the characteristics and causes of effluent quality variations. The treatment plants studied were of the one- and two-stage types, were located in various parts of the country, and ranged in capacity from 0.5 to 33.7 mgd.

Based on the results of this study, the following conclusions have been derived:

1. The distributions of effluent BOD_5 and suspended solids concentrations at the trickling filter plants studied can be described adequately using the log-normal or gamma distributions, or by using a two-parameter empirical distribution model consisting of a series of linear regression equations. The log-normal distribution is preferred because it is generally easier to use than the gamma or empirical distributions.

2. Trickling filter effluent BOD_5 and suspended solids data were not strictly random but contained significant non-random components. Daily suspended solids concentration data were usually more variable and more random than daily BOD_5 concentration data from same plant.

3. Effluent BOD_5 and suspended solids data from just under half of the trickling filter plants studied had significant weekly cycles. Data from different plants tended to have quite different cycles, but overall effluent BOD_5 and suspended solids concentrations tended to be below average on Sundays.

4. Significant non-random variations in monthly average effluent BOD_5 and suspended solids concentrations were found in all 11 plants studied. Normalized data from all plants studied were pooled, and it was found that annual patters existed in both BOD_5 and suspended solids data. Effluent concentrations tended to be higher in the winter and lower in the summer.

5. Variations in effluent BOD_5 and suspended solids concentrations appear to be related to fluctuations in influent and environmental parameters. Using multiple linear regression analysis, it was found that linear combinations of various process parameters could be used to account for between 9 and 91 percent and between 17 and 74 percent of the variation in effluent BOD_5 and suspended solids data, respectively. Influent and primary effluent BOD_5 and suspended solids concentrations and wastewater temperature had the highest correlations with effluent concentrations.

6. A significant percentage of the variation in effluent data from trickling filter plants is caused by the limites of the precision of the BOD_5 and suspended solids tests. It was estimated that between 5 and 78 percent of the variation in the effluent BOD_5 and suspended solids data from the plants studied was caused by measurement error alone.

SECTION 3

LITERATURE REVIEW

Studies of wastewater treatment plant process variability have generally been designed to answer one or more of the following questions:

1. What are the general characteristics of the probability distributions of effluent data and what functions might be used to model these distributions?
2. Are effluent data random in nature or do non-random trends and cycles exist?
3. How might effluent variations be modeled, either in terms of a time series or as a function of independent physical parameters?
4. What are the effects of sampling and measurement errors on effluent data?

In constrast to the numerous studies of process variability in activated sludge plants, relatively few studies of trickling filter effluent variability have been conducted. However, there are similarities in the design, the type of wastes treated, and the factors affecting both types of treatment plants. Consequently, there should be similarities in the general characteristics of their effluent data. Furthermore, the methods used to investigate variability in either process are the same. Hence, research on both types of treatment process is reviewed.

EFFLUENT PROBABILITY DISTRIBUTIONS

Many authors have commented on the distributions of data values for effluent BOD_5 and SS concentrations. (1,17,20,21,32,38,43) Most workers have concluded that effluent concentration data have positive skewness, (1,20,32,38,43) which occurs when there is a larger spread in the data for values above the mean than for values below it. Also, the shape of the distributions for any one parameter varies from plant to plant, and no single distribution function has been found that is in all cases best for modelling the data from all treatment plants. (17,20,38,44) However, in most cases an empirical model (21,39) or the log-normal distribution (20,38,32) can be used to adequately describe effluent distributions.

Pöpel (38) characterized daily data from three activated sludge plants and monthly data from five activated sludge plants and one second stage trickling filter. For daily effluent BOD_5 concentration data, two sets were found to be log-normal, and one set was skew-log-normal. One set of daily SS data was log-normal. The observation was made that for most data, the coefficient of variation was greater

EPA Form 2350-4 (4-80)
(PREVIOUSLY ETN. EPA FORM 337)
TYPING GUIDE SHEET

for the effluent stream than for the influent stream, thus indicating that treatment increased variability.

In a report studying discharge requirements in their area, Sacramento Area Consultants (43) found that effluent BOD_5 concentration data from three activated sludge plants could be described best with gamma distributions.

Hann et al (20) evaluated two years of daily effluent BOD_5 concentration data from three activated sludge and three trickling filter plants. Normal, skew-normal, log-normal, skew-log-normal, Pearson III and log-Pearson III distributions were tested for their fit to each data set. Results varied among different plants and from year to year within each plant, but in general, more of the data could be described with the log-normal and skew-log-normal distributions than with other distributions. The Pearson III distribution was the best fit for data from two trickling filter plants but did not accurately describe data from any other plants. It was also concluded that the coefficient of variation could be used to predict which distribution best described each set of data.

Niku et al (32) examined daily effluent BOD_5 and SS data from 37 activated sludge plants. No single distribution was best in every case for describing treatment plant effluent data, but it was concluded that "the log-normal distribution gives a good overall fit to observed effluent BOD and SS concentrations." The authors found that out of the 37 data sets tested, 24 BOD and 25 SS data sets could be accepted as fitting a log-normal distribution.

Thoman (44) examined daily effluent BOD_5 data from eight treatment plants and found that the distributions "ranged from almost uniform to normal and log-normal (probability) density functions."

Using flow, influent and effluent BOD_5, and influent and effluent SS concentration data from five activated sludge plants, Adams and Gemmel (1) tested the fit of uniform, normal and Poisson distributions. It was concluded that the "data did not significantly fit any of the distributions."

Hovey et al (21) examined descriptive statistics (mean, variance, skew coefficient, etc.) from 28 activated sludge plants and concluded that no single classical probability distribution could be used to describe either BOD_5 or SS effluent concentration data from all plants. However, it was not indicated if specific distributions were actually tested for fit to the data. The authors concluded that a single distribution model to describe data from all plants could be constructed using an empirical relationship developed previously by Porter and Boone, (39) Wheatland (47) and Porter. (40) The model consists of a series of simple linear regression equations which are used to predict precentile values for daily effluent BOD_5 concentrations based on the mean annual concentration. Similar models were developed for daily SS concentrations as well as for 7-day and 30-day continuous averages for both parameters.

RANDOMNESS OF INFLUENT AND EFFLUENT DATA

It has been reported in numerous studies that influent and effluent data are not purely random. Thoman (44), Bertheaux (7,9), Lin and Heinke (24) and others

(1,38) have demonstrated the existence of large non-random components in daily influent and effluent BOD_5 and SS data, as well as flow and wastewater temperature data.

Adams and Gemmel (1) tested data for influent and effluent BOD_5 and SS as well as for flow and found that for all yearly data series, the hypothesis of randomness could be rejected. The data were divided into shorter segments and it was discovered that as the segments were made shorter, an increasing percentage of the series appeared to be random. The authors also determined that the autocorrelation coefficients for most series were much smaller at a lag of 2 days than at a lag of 1 day and concluded that this might imply that composite sampling "tends to disguise the randomness of the variables."

Non-random components in treatment plant data are not unexpected. Factors that cause non-random behavior in the data include the following: cycles in influent strength and flow due to weekly patterns of water use; seasonal precipitation patterns; and seasonal changes in temperature which affect biological activity.

DAILY CYCLES IN EFFLUENT DATA

While the diurnal fluctuations of flow rate and influent concentrations have been recognized for some time, (19,28) no information was found on daily cycles in effluent concentration data.

WEEKLY AND ANNUAL CYCLES IN EFFLUENT DATA

A number of workers have examined average daily flow and effluent concentration data for weekly and annual cyclical patterns. (20,38,45)

Pöpel (38) studied daily data from eight treatment plants and found that some effluent BOD_5 and SS concentration data contained significant weekly cycles while some data sets did not. Monthly average data from eight treatment plants were examined to determine if changes in the data from month to month were statistically significant. Most effluent BOD_5 and all effluent SS and flow data tested varied significantly from month to month. However, no information was given on the actual form of the weekly and annual cycles observed.

Hann et al (55) concluded that effluent BOD_5 concentrations and mass discharges tended to be lower on the weekends at most treatment plants studied; flow rates were generally lower on the weekends at half the plants studied. A seasonal pattern was observed in effluent BOD_5 concentration data from two activated sludge plants and two of four trickling filter plants studied. Effluent BOD_5 concentrations were found to be generally above average in the winter and below average in the summer. No significant yearly pattern was found in flow data.

Townshend (45) studied data from over 300 activated sludge, trickling filter, stabilization pond and septic tank treatment processes in the Province of Ontario. Data were pooled for each process type and tables were prepared listing each months, 15, 50 and 85 percentile value for effluent BOD_5 and SS concentrations. A significant

BEGIN FIRST LINE OF TEXT HERE
CENTER OF PAGE
TOP OF IMAGE AREA
DROPPED HEAD, BEGIN SECTIONS HERE
BEGIN LAST LINE OF TEXT
BOTTOM OF IMAGE AREA OUTSIDE DIMENSION FOR TABLES AND ILLUSTRATIONS
3/8"
PAGE NUMBER
EPA Form 2350-4 (4-80)
(PREVIOUSLY CIN. EPA FORM 287)
U.S. GOVERNMENT PRINTING OFFICE 1980-660-630
TYPING GUIDE SHEET

annual pattern can be seen in the plots of trickling filter data while no significant annual pattern can be observed in the activated sludge data. Monthly average trickling filter BOD_5 concentrations were significantly above average in January through March and significantly below average May through July; SS concentrations were significantly above average February through April and significantly below average June, July, October and November.

STUDIES USING SPECTRAL AND HARMONIC ANALYSIS TO DETECT PERIODIC CYCLES

Spectral and harmonic analysis have been used by a number of workers to detect and model cyclic patterns in treatment plant data. (1,24,44)

Thoman (44), who was first to use spectral analysis to examine wastewater treatment plant data, studied the variability in effluent BOD_5 concentration data from three primary and three secondary plants. Individual results were not presented for each plant, but the author reported that for most plants, a large percentage of the variance in the data was concentrated at the low frequency end of the spectrum and could be modeled by using the first five terms of a Fourier series. A possible conclusion, based on these results, is that annual and seasonal cycles exist in the data. Thoman also found significant weekly cycles in the data from several of the plants studied.

Lin and Heinke (24) used spectral analysis in a study of influent, primary effluent and secondary effluent data from two activated sludge plants. Significant low frequency cycles were evident to BOD_5 and SS water temperature and flow data. Between 24 and 40 percent of the variance in the data could be removed from BOD_5 and SS effluent data by using the first 5 terms of a Fourier series to model the data. With respect to short term cycles, influent and effluent SS data from one plant contained significant weekly cycles; influent BOD_5 and suspended solids data from the other plant exhibited significant weekly cycles. The authors also concluded that primary sedimentation is a purely random process because of the low correlation between imputs and outputs of the process.

Adams and Gemmel (1) used harmonic analysis to check for cyclical variation in influent, effluent and flow rate data from five activated sludge plants. The authors found that, on the average, the first nine terms of a Fourier series could be used to explain approximately 30 percent of the variance in influent and effluent BOD_5 and SS data and approximately 50 percent of the variance in the flow rate data. In most cases, the first term accounted for nearly half the explained variance in the BOD_5 and SS data. The first and fourth terms were predominant for the flow rate data.

Based on the results of the above mentioned studies utilizing spectral and harmonic analysis, it can be concluded that two of the major non-random components of treatment plant effluent data are weekly cycles and long term fluctuations with periods of over 70 days. Their exact nature has not been demonstrated, but the long term fluctuations in effluent quality probably consist of repeating, seasonal changes and non-cylical, long-term fluctuations caused by changes in plant performance.

EPA Form 2350-4 (4-80)
(PREVIOUSLY CIN EPA FORM 347)
TYPING GUIDE SHEET

TIME SERIES AND DETERMINISTIC PHYSICAL MODELS

A number of authors have developed models to examine the effects of process parameters on treatment performance and to investigate the time series behavior of effluent data. (9,14,17,21,24,33)

Berthouex et al (9) used the time series methods of Box and Jenkins (10) to study the relationship between daily effluent BOD_5 concentration and the following parameters: influent BOD_5 concentration, mixed liquor suspended solids concentration, wastewater temperature and values of effluent BOD_5 concentration from previous days. For the treatment plant under study, the authors found that a model using values of effluent BOD_5 concentration from the preceding seven days explained roughly 50 percent of the variance in the data, while a model containing only the independent physical parameters explained less than 15 percent of the variation.

In a study of activated sludge plants treating industrial wastes, Eckenfelder and Roth (17) developed a method to determine the variations in soluble effluent BOD_5 concentration caused by changes in wastewater temperature, mixed liquor suspended solids concentration and cell residence time. Theoretical equations that described the relationships between the different variables were used to determine the effect of each independent parameter. It was assumed that the theoretical relationships between the variables were valid for the data under study. Empirical correlations between the independent variables and effluent BOD_5 were not studied.

Hovey et al (21) used multiple linear regression models to predict secondary effluent BOD_5 and SS concentrations and mass discharges from influent and primary effluent BOD_5 and SS concentrations and flow rate. For the 28 activated sludge plants studied, the authors determined that between 1 and 59 percent of the variance in effluent BOD_5 and SS data could be explained using the regression models. No conclusions were made concerning the effect of influent variation on effluent variation because the authors felt that percentage of variance explainded was generally low and because regression models do not prove cause and effect relationships.

In a study of phosphorus removal, David et al (14) developed a variance components regression model to examine influent-effluent relationships using monthly average data from 12 treatment plants of unspecified type. Significant positive correlations were found between flow and effluent SS, between influent SS and effluent SS and between influent BOD_5 and effluent BOD_5.

Lin and Heinke (24) examined data from two activated sludge plants and found that wastewater temperature is negatively correlated with effluent BOD_5 concentration.

Niku et al (33) examined daily effluent BOD_5 and SS data from 37 activated sludge plants to determine the nature and causes of variation in the data. Descriptive statistics, correlation analysis and regression analysis were used to determine which influent, environmental and process control variables affect effluent concentrations. It was found that no single variable could be used to explain the variability of effluent BOD_5 and SS data from all the plants, and the contribution of any one variable to effluent variation was different from plant to plant. Linear regression models explained from 13 to 61 percent and from 5 to 52 percent of the variation in BOD_5 and SS data, respectively, but the form and the coefficients of the models were different in each plant.

SAMPLING AND MEASUREMENT ERRORS

The effects of sampling and measurement errors on the variability of effluent data and the occurence of discharge violations have been discussed by several authors. (3,7,43)

Bertheaux (7) pointed out that taking Sunday samples will often lower weekly average values and that measurments of effluent concentrations can differ greatly from the true concentrations when only one sample is analyzed. An example is presented of a statistical method for testing the hypothesis that a discharge limit has actually been exceeded when several measurements of the same parameter give different results.

In a report examining discharge requirements and effluent variability at treatment plants in their area, the Sacramento Area Consultants (43) found that a significant percentage of effluent variability at some plants was caused by measurement error alone. The authors concluded that at one point, the number of additional daily effluent concentration violations caused solely by measurement errors was 3 and 9 per year for BOD_5 and SS, respectively.

EPA Form 2350-4 (4-80)
(PREVIOUSLY CIN. EPA FORM 287)
TYPING GUIDE SHEET

BEGIN FIRST LINE OF TEXT HERE
CENTER OF PAGE
TOP OF IMAGE AREA
DROPPED HEAD. BEGIN SECTIONS HERE

SECTION 4

DESCRIPTION OF DATA AND ANALYTICAL METHODS

The questions posed at the beginning of this study have been investigated by statistically analyzing daily performance data from eleven trickling filter treatment plants. The data obtained, the modifications made to the data prior to analysis, and the statistical methods used to examine the data are discussed in this section.

DESCRIPTION OF DATA

The data base for this study consists of daily operation and performance records from eleven trickling filter plants located in six different states. As part of a U.S. Environmental Protection Agency sponsored study (grant #R805097-1), requests for operation records were made to water quality agencies and treatment plant operators in over 30 states. From the replies received, data from eleven plants were chosen for analysis based on the completeness of the data records.

The treatment plants chosen for this study are listed in Table 1. Identification of individual plants is by number and location, rather than by name. Also listed is time period covered by the data, process type and average flow rate. Data sets from most of the plants are one year in length, but the data from plants 5 and 7 are somewhat longer.

Parameters that were requested are used in this study include:

total daily flow, Q
influent BOD$_5$ and suspended solids concentrations, BOD; SS;
primary effluent BOD$_5$ and suspended solids concentrations, BODp, SSp
secondary effluent BOD$_5$ and suspended solids concentrations, BODe, SSe
influent wastewater temperature, WTEMP
average daily air temperature, ATEMP
influent pH, pH;
recycle ratio, R

All concentration data are flow weighted composite values.

Some of the variables listed above were not available from one or more of the treatment plants studied. A list of the parameters made available from each plant is presented in Table 2. Air temperature data for two plants were obtained from U.S. Weather Service records. (11,12,13)

BEGIN LAST LINE OF TEXT
3/8"
BOTTOM OF IMAGE AREA OUTSIDE DIMENSION FOR TABLES AND ILLUSTRATIONS

PAGE NUMBER
EPA Form 2350-4 (4-80)
(PREVIOUSLY CIN. EPA FORM 347)
U.S. GOVERNMENT PRINTING OFFICE
TYPING GUIDE SHEET

TABLE 1

TRICKLING FILTER TREATMENT PLANTS STUDIED

Plant Number	Location	Period of Data	Type of Process	Average Flow (mgd)
1	Iowa	Jan '76 - Dec '76	2-stage	33.7
2	Michigan	Jan '76 - Dec '76	1-stage	0.5
3	Wisconsin	Jan '76 - Dec '76	2-stage	7.6
4	Michigan	Jan '76 - Dec '76	1-stage	0.8
5	Ohio	Feb '76 - July '77	1-stage	15.8
6	Michigan	Jan '77 - Dec '77	2-stage	0.9
7	Michigan	Jan '76 - Jan '77	1-stage	11.6
8	Oklahoma	Jan '76 - Dec '76	2-stage	12.3
9	Michigan	Sept '76 - Aug '77	1-stage	4.0
10	Michigan	Dec '76 - Nov '77	1-stage	0.8
11	Arkansas	Jan '77 - Dec '77	2-stage	6.0

TABLE 2

DAILY DATA AVAILABLE FROM EACH PLANT

Plant Number	Q	BOD_i, SS_i	BOD_p, SS_p	BOD_e, SS_e	WTEMP	pH	ATEMP	R
1	x	x		x	x	x	x[b]	x
2	x	x	x	x	x	x		
3	x	x	x	x	x	x	x	
4	x	x[a]	x[a]	x[a]	x			
5	x	x	x	x	x	x		x
6	x	x	x	x	x			
7	x	x	x	x	x	x	x[b]	
8	x	x		x	x			
9	x	x	x	x				
10	x	x[a]	x[a]	x[a]	x			x
11	x	x	x	x				

[a] Data for five days per week.

[b] Data obtained from U.S. Weather Service records.

EPA Form 2350-4 (4-80)
(PREVIOUSLY CIN. EPA FORM 287)
TYPING GUIDE SHEET

BEGIN FIRST LINE OF TEXT HERE
CENTER OF PAGE
TOP OF IMAGE AREA

TREATMENT OF MISSING DATA

Statistical analyses ideally are performed on data sets which have no missing values. Unfortunately, gaps of various duration do exist in the data for most parameters from almost all of the treatment plants studied. For some data, the number of missing cases is a significant percentage of the total record. For example, the BOD$_5$ and SS data sets from plants number 4 and 10 contain over 110 missing values each because these parameters were not sampled on weekends. However, for most of the plants in this study, the number of missing cases for all variables is less than five percent.

DROPPED HEAD; BEGIN SECTIONS HERE

Different procedures for handling missing data are used with different statistical methods. To meet the missing data requirements of the statistical methods used in this study, two separate data files were prepared.

The first data file was prepared by inserting the missing value label "-99", for each case where data were unavailable. Each "labeled" data series was identical to the original treatment plant data except that missing data cases were labeled, and therefore data series of different parameters from the same plant were equal in length.

A "reduced" data file was prepared from the labeled file by deleting those cases containing missing value labels. Thus, for each labeled data series, a new series was created which had no missing values, but in general, contained fewer cases than its corresponding labeled series. Also, because different variables had different numbers of missing values, the length of the data series for different variables from the same plant were generally of different lengths.

A list of statistical procedures used in this study and the corresponding data file used with each procedure appears in Table 3.

The exact numerical results of analyses performed on data sets containing missing values will be different from what they would be if data were available for every case. However, for most data series used in this study, the number of missing values is relatively small. Therefore, it is unlikely that the results of any of the statistical tests performed were affected substantially because the data were not totally complete.

ANALYTICAL METHODS

The methods used to investigate the questions asked at the beginning of this study are listed in Table 3 and outlined in this section. A brief introduction to some of the lesser known statistical methods used will be presented, but it will be assumed that the reader is familiar with the more common techniques.

Statistical computations for this study were performed on the Burrough's 6700 computer at the University of California, Davis Computer Center. Some of the computer programs used to analyze the data were obtained from Statistical Package for the Social Sciences. (31)

BEGIN LAST LINE OF TEXT
BOTTOM OF IMAGE AREA OUTSIDE DIMENSION FOR TABLES AND ILLUSTRATIONS
3/8"

PAGE NUMBER
EPA Form 2350-4 (4-80)
(PREVIOUSLY CIN. EPA FORM 287)
TYPING GUIDE SHEET

BEGIN FIRST LINE OF TEXT HERE
CENTER OF PAGE
TOP OF IMAGE AREA
DROPPED HEAD: BEGIN SECTIONS HERE

TABLE 3

STATISTICAL ANALYSES PERFORMED AND CORRESPONDING DATA FILES

Statistical Method	Type of Data File
Summary Statistics	Reduced
Linear Regression for Computing Empirical Distribution Model Coefficients	Reduced
Kolmogorov-Smirnov Goodness-of-Fit Test	Reduced
Runs Test	Reduced
Serial Correlation	Labled
Analysis of Variance	Labled
Multiple Linear Regression for Physical Models	Labled

BEGIN LAST LINE OF TEXT
BOTTOM OF IMAGE AREA: OUTSIDE DIMENSION FOR TABLES AND ILLUSTRATIONS
3/8"

PAGE NUMBER
EPA Form 2350-4 (4-80)
(PREVIOUSLY CIN: EPA FORM 287)
U.S. GOVERNMENT PRINTING OFFICE: 1980 [illegible]
TYPING GUIDE SHEET

BEGIN FIRST LINE OF TEXT HERE
CENTER OF PAGE
TOP OF IMAGE AREA
DROPPED HEAD, BEGIN SECTIONS HERE

Characterization of Effluent Distributions

The first question asked as part of the objectives of this study was: "How can the distribution of effluent BOD_5 and SS concentrations from trickling filter plants be characterized?" The distributions of effluent data are important because they can be used as convenient summaries of the long-term performance of existing treatment plants and can be used to help predict effluent concentrations for treatment plants under design. Also, distribution models are used in water quality simulations to construct algorithms, used to generate synthetic treatment plant discharge data.

To characterize the effluent BOD_5 and SS distributions for the treatment plants in this study, summary statistics, such as the mean, standard deviation and skew coefficient, were computed for daily, 7-day average and 30-day average effluent concentrations from each plant. Also, three theoretical and two empirical distribution functions were chosen as possible models for effluent distributions and were tested using the Kolmogorov-Simirnov goodness-of-fit test. Details of these procedure are explained below.

Running Averages—

Portions of federal and state discharge requirements for BOD_5 and SS concentrations are stated in terms of running seven and thirty days averages. Therefore, the distributions of these averages, as well as the distributions of daily averages, are of interest and will be investigated in this study.

If trickling filter effluent data were made up of independent, randomly generated series of numbers, it would be possible, using the central limit theorem of statistics, to prove that the distribution of the running average for a variable would assymtotically approach a normal distribution as the length of the average became large. In fact, if effluent data were random, the normal distribution would be a good approximation for averages as small as five days. (27)

As will be demonstrated, trickling filter effluent data are not random, and it cannot be assumed that running average distributions are approximately normal. Therefore, the distributions of running averages had to be determined in the same manner as the distributions of daily averages.

Running 7-day and 30-day averages were computed for effluent BOD_5 concentration data according to the followng formula:

$$R7DB_i = \frac{1}{m} \sum_{k=i-6}^{i} BOD_{e,i} \qquad i = 7,8,9,\ldots n \qquad (1)$$

$$R30DB_i = \frac{1}{m} \sum_{k=i-29}^{i} BOD_{e,i} \qquad i = 30,31,32,\ldots n \qquad (2)$$

where $R7DB_i$ = running 7-day average effluent BOD_5 for day i

$R30DB_i$ = running 30-day average effluent BOD_5 for day i

BEGIN LAST LINE OF TEXT
BOTTOM OF IMAGE AREA OUTSIDE DIMENSION FOR TABLES AND ILLUSTRATIONS

PAGE NUMBER
EPA Form 2350-4 (4-80)
(PREVIOUSLY CIN. EPA FORM 287)
U.S. GOVERNMENT PRINTING OFFICE
TYPING GUIDE SHEET

BEGIN FIRST LINE OF TEXT HERE
CENTER OF PAGE
TOP OF IMAGE AREA
DROPPED HEAD; BEGIN SECTIONS HERE

m = 7-j for running 7-day averages

m = 30-j for running 30-day averages

j = number of missing data from day i-6 (or i-29) through day i

$BOD_{e,i}$ = effluent BOD_5 on day i

Running 7-day and 30-day averages for effluent SS concentration data were computed in a similar manner.

Summary Statistics—

The following summary statistics have been computed for daily, 7-day average and 30-day average effluent BOD_5 and SS concentration data from each plant:

arithmetic mean
standard deviation
coefficient of variation
skew coefficient
maximum observation
minimum observation
number of valid observations

Examination of these statistics will provide information on the central tendencies, ranges and shapes of the effluent BOD_5 and SS concentration distributions for data from each plant.

Theoretical Distribution Models—

Virtually an unlimited number of theoretical distribution functions could be tested for their fit to each set of effluent data. However, practical limitations prevent testing more than a limited number. The normal, log-normal and gamma distribution functions have been chosen for testing because of their relative simplicity, their widespread use, and because their usefulness in modeling treatment plant effluent data distributions has been demonstrated in previous work. (38,43,44) Details of the procedure used to test the fit of each distribution function to trickling filter effluent data will be presented after two empirical distribution models are discussed.

Empirical Distribution Models—

Several authors have demonstrated that there are strong linear relationships between the mean value and each percentile value in treatment plant effluent concentration data. (21,39,47) For example, if series of one year's effluent BOD_5 concentration data from a number of treatment plants are each placed in ascending order and assigned percentile rankings, the 95th percentile value for each plant will be approximately proportional to that plants average annual effluent BOD_5 concentration.

Using this relationship between the mean and percentile values, empirical models for the distribution of effluent concentration data have been developed. (21,40)

BEGIN LAST LINE OF TEXT
BOTTOM OF IMAGE AREA OUTSIDE DIMENSION FOR TABLES AND ILLUSTRATIONS

PAGE NUMBER
EPA Form 2350-4 (4-80)
(PREVIOUSLY CIN. EPA FORM 357)
U.S. GOVERNMENT PRINTING OFFICE
TYPING GUIDE SHEET

The following linear regression models, used by Hovey et al (21) to predict effluent concentrations in activated sludge plants, are used in this study to model trickling filter plant effluent data distributions:

$$X\%\text{ile}\ (BOD_e) = A_x \cdot \overline{BOD_e} + C_x \qquad (3)$$

$$X\%\text{ile}\ (SS_e) = A_x \cdot \overline{SS_e} + C_x \qquad (4)$$

$$X\%\text{ile}\ (R7DB) = A_x \cdot \overline{R7DB} + C_x \qquad (5)$$

$$X\%\text{ile}\ (R7DS) = A_x \cdot \overline{R7DS} + C_x \qquad (6)$$

$$X\%\text{ile}\ (R30DB) = A_x \cdot \overline{R30DB} + C_x \qquad (7)$$

$$X\%\text{ile}\ (R30DS) = A_x \cdot \overline{R30DS} + C_x \qquad (8)$$

where X%ile (y) = value of y (g/m^3) which is greater than X percent of the data values

$\overline{y}$ = arithmetic mean of all values of y

A_x, C_x = regression coefficients, different for each model

x = 0, 10, 20, ..., 90 and 95, 98, 99, 100

y = BOD_e, SS_e, R7DB, R7DS, R30DB, R30DS

Coefficients for each of the above models are estimated from performance data from a representative group of treatment plants. The data from each plant is first ranked in ascending order and percentile values are determined for each value of X. Then, for each value of X, the coefficients A_x and C_x are estimated by performing a simple linear regression using the mean and Xth percentile value from each plant as one case. (Thus, if data from five treatment plants are used, there will be five x,y pairs for a regression of the form $y = A_x + C$).

The set of linear regression equations from any of the above models do not by themselves define a continuous probability distribution. However, given the mean value from a set of data, a finite number of predicted percentile values can be determined. A continuous distribution can then be constructed by linearly interpolating between each of the predicted percentile values.

The models defined by equations 3 through 8 all have the same form and only differ in the values for the coefficients A_x and C_x. Also, all six models are one-parameter distribution models because only one quantity (the mean value) is required to calculate the estimated distribution for a given data set. It would be expected that an empirical model in which an additional summary statistic from the data set is utilized (e.g. the standard deviation), would result in a better estimate of the actual distribution. Such two parameters distribution models are quite common, and in fact, the three theoretical models used in this study are all two parameter models. It was therefore decided to develop a set of two parameter empirical models similar to equations 3 through 8, but using the mean and standard deviation as

EPA Form 2350-4 (4-80)
(PREVIOUSLY CIN. EPA FORM 287)
TYPING GUIDE SHEET

independent parameters instead of only the mean. The two-parameter models are defined as follows:

$$\text{X\%ile}\ (\text{BOD}_e) = A_x \cdot \overline{\text{BOD}}_e + B_x \cdot \text{SD}\ (\text{BOD}_e) + C_x \qquad (9)$$

$$\text{X\%ile}\ (\text{SS}_e) = A_x \cdot \overline{\text{SS}_e} + B_x \cdot \text{SD}\ (\text{SS}_e) + C_x \qquad (10)$$

$$\text{X\%ile}\ (\text{R7DB}) = A_x \cdot \overline{\text{R7DB}} + B_x \cdot \text{SD}\ (\text{R7DB}) + C_x \qquad (11)$$

$$\text{X\%ile}\ (\text{R7DS}) = A_x \cdot \overline{\text{R7DS}} + B_x \cdot \text{SD}\ (\text{R7DS}) + C_x \qquad (12)$$

$$\text{X\%ile}\ (\text{R30DB}) = A_x \cdot \overline{\text{R30DB}} + B_x \cdot \text{SD}\ (\text{R30DB}) + C_x \qquad (13)$$

$$\text{X\%ile}\ (\text{R30DS}) = A_x \cdot \overline{\text{R30DS}} + B_x \cdot \text{SD}\ (\text{R30DS}) + C_x \qquad (14)$$

where A_x, B_x, C_x = regression coefficients, different for each model

x = 0, 10, 20, ..., 90 and 95, 98, 99, 100

Coefficients for the models defined by equations 9 through 14 are determined in the same manner as coefficients for the one-parameter models except that the mean and standard deviation from each plant are used as independent variables for the regression.

Kolmogorov-Smirnov Goodness-of-Fit Test—

The usefulness of each of the five (proposed) distribution functions for modelling BOD_5 and SS data from each plant was examined using the Kolmogorov-Smirnov goodness-of-fit test. (5) The null hypothesis for this test is:

> "The observed data were sampled from an underlying population which is distributed according to the hypothesized distribution function."

It is assumed that if the test hypothesis is true, any differences between the observed distribution and the hypothesized distribution are due to the effects of random sampling.

To perform the K-S test, the observed data are ranked in ascending order and assigned percentile values. The difference, d_i between the observed percentile and the percentile predicted by the hypothesized distribution is then computed for each data value. The maximum absolute value of d_i is the test statistic, D. A graphical illustration of the computation of D is presented in Figure 3. The probability, P, that the test hypothesis is true is computed from the value of D and the number of data points, n. (5)

The K-S test is especially useful for comparing a hypothesized distribution with the distribution of a set of data points because the test statistic, D, provides a quantitative measure of the actual percentile difference between the two distributions. The value of D can be thought of as the maximum absolute error that would occur if the hypothesized distribution were used to estimate the percentiles of the observed data values.

EPA Form 2350-4 (4-80)
(PREVIOUSLY CIN: EPA FORM 287)
TYPING GUIDE SHEET

BEGIN FIRST LINE OF TEXT HERE
CENTER OF PAGE
TOP OF IMAGE AREA
DROPPED HEAD: BEGIN SECTIONS HERE

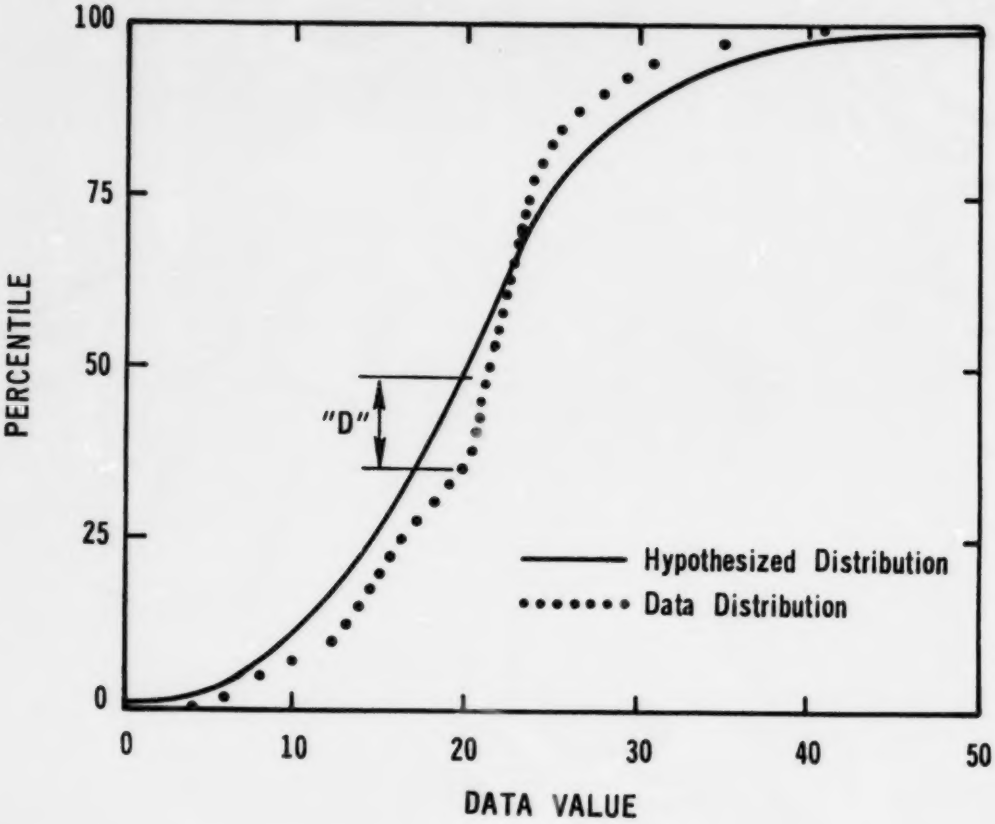


Figure 3. Example of Kolmogorov-Smirnov goodness of fit test.

BEGIN LAST LINE OF TEXT
3/8"
BOTTOM OF IMAGE AREA OUTSIDE DIMENSION FOR TABLES AND ILLUSTRATIONS

PAGE NUMBER
EPA Form 2350-4 (4-80)
(PREVIOUSLY CIN. EPA FORM 287)
TYPING GUIDE SHEET

From experience with the K-S test, it can be shown that the largest values of d_i tend to occur between the 30th and 70th percentiles and that the percentile differences between the two distributions are generally smaller in the tails. Therefore, the value of D is typically much larger than the values of d_i for data points in the extremes of the observed distribution.

Randomness of Effluent Data—

The second question asked as part of the objectives of this study was: "Are trickling filter effluent BOD_5 and suspended solids concentration data random?" If trickling filter effluent data are not random, then distribution models for effluent concentrations cannot be used in the classical manner because probability density functions are usually assumed to be based on random data. Another consequence of non-randomness would be that predictable cycles might exist in effluent data.

The randomness of trickling filter effluent BOD_5 and SS concentration data have been tested using the runs test and a test for serial correlation. In addition to simply determining if the data are random, these tests can be used to determine the relative magnitude of non-randomness in each set of data and to provide a basis for comparing the relative randomness of different sets of data.

Runs Test—

The runs test is a general purpose test of the randomness of a data series. Trend, grouping and many types of cyclical movement in the data can be detected. The test hypothesis for this test is:

The data series have been generated by a purely random process.

A "run" is defined as a maximal consecutive sequence of data points such that every point in the sequence is greater than the median value of the data series or such that every point is less than the median value of the data series. "Maximal" sequence means that if a sequence of data points meets the definition of a run, then no subset of that sequence is considered a run. For example, the series of data points plotted in Figure 4 contains five runs.

The probability that the test hypothesis is true (i.e., the data are random) is determined by first computing the quantity Z, as follows:

$$Z = \frac{2R - (N + 2)}{\left[\frac{N^2 - 2N}{N-1}\right]^{1/2}} \qquad (15)$$

where R = number of runs

N = n, if n even
n-1, if n odd

n = number of data values

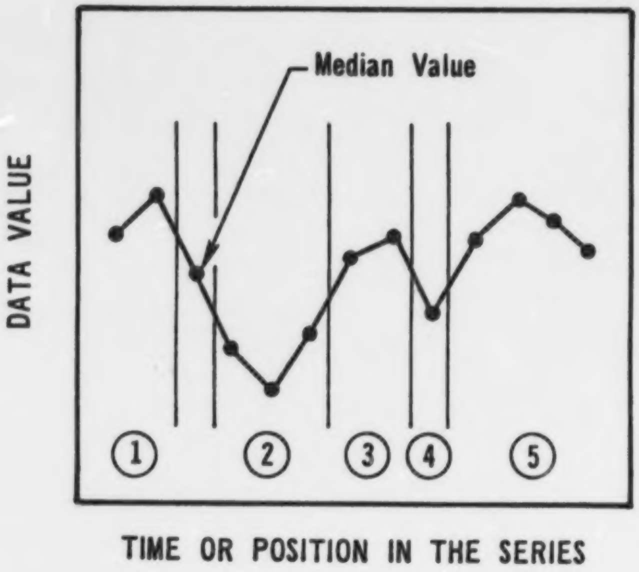


Figure 4. Example of counting runs in a data series.

The statistic Z is a z-score from a standard normal distribution (normal distribution with a mean of zero and a standard deviation of one). The probability that the data set is random is twice the area under a standard normal density curve to the right of the value of Z. Normal curve areas corresponding to various z-scores are given in tables in most statistics tests. (27)

Serial Correlation—

In some cases, the runs test fails to detect dependence between consecutive data pairs (serial correlation) in addition to the runs test when examining data randomness.

The test for serial correlation is performed by treating the original data series as one variable, x_t, and treating the data series shifted forward one observation as the other variable, x_{t-1}. The Pearson correlation coefficient[a] a measure of associative of the strength of the linear relationship between two variables, is then computed to test the correlation between x_i and x_{i-1}. If the data are random, the correlation coefficient will tend toward zero.

Annual and Weekly Cycles

The third question asked as part of the objectives of this study was: "If trickling filter plant effuent BOD_5 and SS data are not random, do the data follow any annual or weekly cycles?" If effluent data are not random, then it will be helpful to know if there are any predictable cycles in treatment plant performance. Knowledge of these cycles would provide information on the dynamics of treatment processes and would be useful in predicting process performance.

To test for the possibility of weekly or annual cycles in trickling filter effluent data, a two-way analysis of variance was performed on effluent BOD_5 and suspended solids concentration data from each plant. Analysis of variance was not used to explicitly determine if cycles existed in the data series, but merely to determine if effluent concentrations were dependent on the month or day of the week on which they occurred. However, if effluent BOD_5 (or SS) concentrations from a treatment plant were found to be dependent on the day of the week, it was concluded that a weekly cycle existed in that data set.

A similar conclusion concerning annual cycles could not be made for data that were found to be dependent on the month. With only one year of data from each plant, there was no way of determining if an observed annual sequence was representative of a repeating annual cycle or merely part of a random fluctuation in average monthly performance. All that could be concluded in this case was that some variation in the data was due to significant changes in average performance from month to month.

---

[a] The Pearson correlation coefficient is defined as:

$$r_{x,y} = \frac{1}{n} \sum_{i=1}^{n} \frac{x_i - \bar{x}}{S_x} \frac{y_i - \bar{y}}{S_y}$$

The exact form of weekly patterns and changes in monthly averages has not been presented for each individual treatment plant because such detailed information would be time consuming to interpret and would tend to obscure any general patterns common to all treatment plants.

Effluent data from all plants were pooled to test for the existence of annual or weekly cycles common to all trickling filter plants. First, effluent BOD_5 (and SS) data from each plant were normalized by subtracting the series mean from each data value and dividing the result by the standard deviation. Data from all plants were then pooled together. Plots of the group averages for data grouped by month and for data grouped by day of the week were prepared and examined for significant patterns.

Analysis of Variance—

Analysis of variance, (ANOVA) is a statistical method for examining the effects of one or more qualitative independent variables on the response of a quantitative dependent variable. It is used in this study to determine if effluent BOD_5 and SS concentrations at individual plants are dependent on the month or day of the week on which they occur. While a number of other questions can be answered using ANOVA methods, a detailed study of weekly and monthly cycles in effluent data from each individual plant is beyond the scope of this study.

In analysis of variance, the variation in the response of the dependent variable is assumed to be caused by random error and (possibly) by variation in the independent variables. To test for the existence of a relationship between an independent variable and the dependent variable, data for the dependent variable are grouped according to the corresponding value of the independent variable for each observation. (For example, to test for the effect of day of the week, effluent BOD_5 concentration data would be separated into seven groups, each one corresponding to the day of the week on which the BOD sample was taken).

If the independent variable has no effect on the dependent variable, then the averages of all the groups will tend to be equal. However, there will always be some differences between group averages because of random variation in the dependent variable. If, on the other hand, the response of the dependent variable is affected by the value of the independent variable, then the differences between group averages will be greater than in the case of no effect.

A common statistical procedure, called an F-test, is used to test for statistically significant differences in group averages.

In the analysis of variance, a linear relationships is assumed between the independent variables and the dependent variable. For the analyses performed in this study, the following model was assumed:

$$y = \mu + \alpha_i + \beta_j + \gamma_{ij} + \varepsilon \qquad (16)$$

where y = dependent variable, either BOD_e or SS_e

μ = mean value of y

α_i = effect of ith month on the value of y

β_j = effect of jth day of the week on the value of y

γ_{ij} = interaction effect of the ith month and the jth day of the week on the value of y

ε = random error, assumed to normally distributed with mean zero.

The γ term is necessary because of the additive form of the model. Relationships may exist between the dependent and independent variables that cannot be explained by a model of the form

$$y = \mu + \alpha + \beta + \varepsilon \tag{17}$$

For example, the effects of month and day of the week may not be strictly additive, or the effect of day of the week may change from month to month. The exact nature of significant interaction effects has not been studied, but if they exist, it can be concluded that the effects of month and day of the week are not simply additive, that is, not of the form shown in equation 17.

Significance level, x, which is the probability of rejecting a hypothesis that is in fact true is used as a test criteria. Rejecting a correct hypothesis is called a type I error. For most of the statistical procedures in this study, a 5 percent significance level has been used as the criterion for rejecting test hypothesis. However, a computed significance of 1 percent has been used when testing for the effects of day-of-week and month because some of the assumptions of the F-test are not met by the treatment plant data used in this study. Norton (34) has shown that if the assumptions upon which the F-test is based are approximately true, the test can still be used, but the computed probability that the test hypothesis is true is actually lower than the true probability. For example, if the computed probability that there is no effect of month is 2 percent, the actual probability might be 5 percent. Norton's work shows that for data similar to the data used in this study, a computed probability of 1 percent corresponds to an actual probability of 3 to 5 percent.

Because the existence of interaction effects is an unfavorable result, a computed probability of 5 percent will be used as the rejection criterion for the hypothesis of no interaction effects.

Effect of Various Process Parameters

The fourth question asked in the beginning as part of the objective was: "What are the effects of influent, environmental and operational parameters on effluent variability?" This question has been examined by using multiple linear regression analysis to study the correlations between various process parameters and effluent BOD_5 and SS concentrations.

The analyses discussed in the preceeding sections have been designed to examine the fluctuations in effluent BOD_5 and SS data without consideration of the causes of these variations. In addition to understanding how trickling filter performance varies it will be helpful to understand why it varies. In this section and the next,

EPA Form 2350-4 (4-80)
(PREVIOUSLY CIN. EPA FORM 287)

TYPING GUIDE SHEET

the procedures used to examine some causes of trickling filter plant effluent variability are discussed.

The regression analysis has been conducted in two parts. First, multiple linear regression equations predicting effluent BOD_5 and SS concentrations for each plant were constructed using linear combinations of the following independent variables:

flow, Q

influent BOD_5 and suspended solids concentrations, BOD_i, SS_i

primary effluent BOD_5 and suspended solids concentrations, BOD_p, SS_p

the inverse of the average influent water temperature, $WTEMP^{-1}$

the inverse of the average air temperature $ATEMP^{-1}$

the deviation from 7.0 of the influent pH, 7.0-pH

recycle ratio, R

Second, to test the relative importance of quadratic terms, loading factors, combinations of terms and values of parameters from previous days (logged variables), regression equations using the terms above as well as the following terms were constructed:

quadratic terms: BOD_i^2, SS_i^2, BOD_p^2, SS_p^2

loading factors: $Q \cdot BOD_i$, $Q \cdot SS_i$, $Q \cdot BOD_p$, $Q \cdot SS_p$

combinations of terms: $T_{avg} = (ATEMP + WTEMP)/2$, $T_d = |ATEMP - WTEMP|$

lagged terms: $BOD_{i,t-1}$, $SS_{i,t-1}$, $BOD_{p,t-1}$, $SS_{p,t-1}$, $BOD_{i,t-2}$, $SS_{i,t2}$, $BOD_{p,t-2}$, $SS_{p,t-2}$, $ATEMP_{t-1}$

In each part, only those terms which were statistically significant and contributed materially to the predictive ability of the regression equation were actually included in the final form of each regression equation.

Multiple Linear Regression—

Simple linear regression is a method for studying the linear relationship between one independent variable and one dependent variable. Multiple linear regression is an extension of simple regression to the case where there is more than one independent variable.

The following model is used in a multiple regression analysis:

$$y = a_1x_1 + a_2x_2 + \ldots + a_nx_n + C + \varepsilon \qquad (18)$$

where y = dependent variable (BOD_e, SS_e)

$x_1, x_2, \ldots x_n$ = independent variables (all other parameter data taken)

$a_1, a_2 \ldots a_n$ = regression coefficients

C = constant

ε = random error, assumed to be from a normal distribution with mean value of zero.

n = number of independent variables

Coefficient of Determination—

The coefficient of determination, r^2, is a measure of the fraction of the variation in the independent variable that can be "explained" with the regression equation. It is defined as:

$$r^2 = \frac{(\text{variation in } Y) - (\text{variation in } Y - \hat{Y})}{(\text{variation in } Y)} = \frac{S_y^2 - S_r^2}{S_y^2} \qquad (19)$$

where S_y^2 = variance of the dependent variable

S_r^2 = variance of residuals, r

$r = y - \hat{y}$

y = dependent variable

$\hat{y}$ = value of y predicted by the regression equation.

The value of r^2 can be between 0 and 1 inclusive. If $r^2 = 1$, there is a perfect linear correlation between the combination of independent variables and the dependent variable. If $r^2 = 0$, the independent variables are uncorrelated with the dependent variable.

Stepwise Inclusion of Variables—

In multiple regression, it is often desirable to include in the regression equation only those variables that significantly contribute to the predictive ability of the equation. Stepwise inclusion of variables has been used in this study to choose which variables are to be included in the final form of each regression equation. In this procedure, variables are entered into the regression equation one at a time. At each step, the variable which causes the greatest increase in the value of the coefficient of determination is entered into the equation. Variables are added to the equation

BEGIN FIRST LINE OF TEXT HERE
CENTER OF PAGE
TOP OF IMAGE AREA

until there are no additional variables whose inclusion would cause a significant increase in the coefficient of determination for the regression equation. The significance of changes in the coefficient of determination, r^2, is determined with a standard F-test.

DROPPED HEAD: BEGIN SECTIONS HERE

F Statistic—

The F-test is used in regression analysis to test the statistical significance of individual variables, groups of variables and the entire regression equation. The null hypothesis for the F-test is:

> "The observed change in r^2 caused by adding a variable (or variables) to the regression equation was caused by the effects of random sampling."

Due to the nature of the data used in this study, it is doubtful that the underlying assumptions of the F-test are strictly met. However, as was mentioned previously, the F-test is reasonably resistant to moderate violations of the assumptions upon which it is based.

The primary use of the F-test in this regression analysis is for deciding at which point in the stepwise inclusion procedure that no additional variables shall be entered into a regression equation. Even if the statistical significance of values for the F-statistic are changed somewhat, the F-test can still be used as a logical criterion for choosing which variables are to be included in each equation.

Missing Data

Listwise deletion has been used to handle missing data. When data for any variable was missing for a particular day, all variables for that day were considered missing and that case was dropped from the analysis. While a considerable amount of data is sometimes lost when this method is used, other methods for handling missing data can cause mathematical anomalies in the results of the regression analysis that make the interpretation of those results difficult.

Normalized Data—

The coefficients of the independent variables in a regression equation can be difficult to interpret when different variables have been measured on widely different numerical scales. For example, consider the following regression equation:

$$x = 415\ y + 0.0211z - 15.0 \qquad (20)$$

Without additional information, it might appear that y has a larger effect than z on the dependent variable x. However, the large coefficient for y might be due to the fact that the average value for y is much smaller than the average value for z and not because y actually has a larger effect on x than z.

To overcome this difficulty, the data for variables used in the regression analyses were normalized by subtracting the mean for the variable from each data value and then dividing the result by the standard deviation. For every variable, the resulting data series had a mean of zero and a standard deviation of one. The coefficients of the resulting regression equations are different from what they would

BEGIN LAST LINE OF TEXT
BOTTOM OF IMAGE AREA OUTSIDE DIMENSION FOR TABLES AND ILLUSTRATIONS

PAGE NUMBER
EPA Form 2350-4 (4-80)
(PREVIOUSLY CIN. EPA FORM 287)
TYPING GUIDE SHEET

TABLE 4

REPORTED PRECISION OF BOD_5 AND SUSPENDED SOLIDS TESTS

Test and Mean g/m^3	Standard Deviation g/m^3	Standard Deviation % of Mean	Source
BOD_5:			
175	26	15%	Standard Methods[42]
175	26	15%	EPA[29]
2.1	0.7	33%	EPA
Suspended Solids:			
1707	13	0.8%	Standard Methods
242	24	10%	Standard Methods
15	5.2	35%	Standard Methods

BEGIN FIRST LINE OF TEXT HERE
CENTER OF PAGE
TOP OF IMAGE AREA
DROPPED HEAD: BEGIN SECTIONS HERE

TABLE 5

ESTIMATED PRECISION OF BOD$_5$ AND SUSPENDED SOLIDS TESTS FOR MEAN CONCENTRATIONS BETWEEN PUBLISHED VALUES [29, 42]

Test and Mean, g/m^3	Standard Deviation as Fraction of the Mean	Standard Deviation as Fraction of the Mean for Two Replicates
BOD$_5$:		
10	0.25	0.18
17	0.23	0.16
21	0.22	0.16
28	0.21	0.15
36	0.20	0.14
48	0.19	0.13
66	0.18	0.13
Suspended Solids:		
15	0.35	
17	0.33	
21	0.30	
27	0.27	
32	0.25	
35	0.24	
38	0.23	
42	0.22	
47	0.21	
52	0.20	

BEGIN LAST LINE OF TEXT
BOTTOM OF IMAGE AREA: OUTSIDE DIMENSION FOR TABLES AND ILLUSTRATIONS

PAGE NUMBER
EPA Form 2350-4 (4-80)
TYPING GUIDE SHEET

be if non-normalized data were used. The regression equation shown above (20) might now have the form:

$$x = 0.091\ y + 0.25\ z \tag{21}$$

The coefficients of a regression equation computed from normalized data can be interpreted as follows: If all other independent variables are held constant, an increase of one standard deviation in the value of an independent variable will cause a change in the independent variable equal to one standard deviation times the value of the coefficient of the independent variable.

Let χ_m = measured value of BOD_5 (or suspended solids)

χ_s = true value of BOD_5 (or suspended solids) in the sample

ϵ_a = analytical measurement error

The quantity ϵ_a is an independent random variable with mean zero.

Because ϵ_a and χ_m are independent, the variance of χ_m is

$$V(\chi_m) = V(\chi_s) + V(\epsilon_a) \tag{22}$$

The estimated percentage of the variance (EPV) of χ_m due to analytical error is:

$$EPV = \frac{V(\epsilon_a)}{V(\chi_m)} \times 100 \tag{23}$$

This quantity is analagous to the coefficient of determination, r^2, which is the fraction of the total variation in the dependent variable explained by a regression equation.

The value of $V(\epsilon_a)$ can be derived from information on the precision of the BOD_5 and suspended solids test. (29,42) Estimates of the precision of these two tests are present in Table 4. For both tests, the standard deviation of the test error, expressed as a fraction of the mean value, appears to be a function of the mean value of the sample. (It was assumed in these studies that the mean value was the true concentration in the sample). Estimates of the standard deviation of the BOD_5 and suspended solids tests for sample values in between those given in Table 4 were determined by plotting the means and fractional standard deviations for each test on log-log paper and linearly interpolating between points. Estimates of the standard deviation and fractional standard deviation for various sample mean values in between those given in the literature (29,42) are shown in Table 5.

When more than one measurement is made of a constituent in a sample, the standard deviation of the average of the measurements will be less than the standard deviation of single measurements. Assuming all measurements are independent and random, the standard deviation of the average of the measurements will be less than the standard deviation of single measurements. Assuming all measurements are

EPA Form 2350-4 (4-80)

independent and random, the standard deviation of the average of replicate measurements made of the sample will be

$$SD(\bar{\chi}_n) = \frac{SD(\chi)}{n} \tag{24}$$

where $SD(\bar{\chi}_n)$ = standard deviation of the mean of n measurements of χ

$SD(\chi)$ = standard deviation of one measurement of χ

When the regression equation is computed from normalized data, the constant of the regression equation will always be zero.

Measurement Error

The final question asked as part of the objectives of this study was: "What is the effect of measurement error on effluent data variability?" Any discussion of the causes of variation in a series of measured quantities should include consideration of the effect of measurement error. Like all laboratory procedures, the BOD_5 and SS tests have a finite degree of precision. Thus, even if the effluent from a treatment plant was of perfectly uniform quality at all times, and in addition, the effluent was always sampled without error, the measured BOD_5 and SS concentrations would still vary from sample to sample because of the inherent variability in the methods of anlaysis.

It is possible to determine how much variation in trickling filter effluent data is caused by measurement errors because the variation due to the limited precision of the BOD_5 and SS tests is independent of other sources of variation in the data. Consequently, for each sample measurement,

$$\chi_m = \chi_s + \varepsilon_a \tag{25}$$

The number of replicates used for BOD_5 and SS measurements was not known for the treatment plants in this study. For the purposes of estimating the variance caused by measurement error, it has been assumed that 2 replicates were used in BOD_5 determinations and 1 replicate was used in suspended solids determinations. The estimated standard deviations of the mean of two replicates is given in Table 5 for the BOD_5 test.

One additional simplifying assumption has been made to estimate the variation due to measurement error. Because the precision of the BOD_5 and SS tests are a function of the concentrations being measured, the precision of these test varies from day to day. For simplicity, it will be assumed that the fractional standard deviations of the measurement errors are a function of each plants annual mean effluent BOD_5 and SS concentrations.

The standard deviation of the measurement errors is now given by

$$SD(\varepsilon_a) = C \cdot \bar{\chi} \tag{26}$$

EPA Form 2350-4 (4-80)
(PREVIOUSLY CIN. EPA FORM 237)
TYPING GUIDE SHEET

where C = estimated standard deviation of the measurement errors, expressed as a fraction of the mean of the data set (values of C from Table 5)

$\bar{\chi}_m$ = annual mean BOD_e (or SS_e)

The quantity EPV is given by

$$EPV = \frac{V(\varepsilon_a)}{V(\chi_m)} \tag{27}$$

$$= \frac{[SD(\varepsilon_a)]^2}{[SD(\chi_m)]^2} \tag{28}$$

$$= \frac{(C \cdot \bar{\chi}_m)^2}{[SD(\chi_m)]^2} \tag{29}$$

$$= \frac{C^2}{[CV(\chi_m)]^2} \tag{30}$$

where $CV(\chi_m)$ = coefficient of variation of χ_m.

SECTION 5

RESULTS

The statistical procedures discussed in the previous section have been used to obtain answers to the questions concerning the variability of trickling filter plant effluent that were posed at the beginning of this study. The results of these analyses are presented in this section.

Empirical and theoretical probability distribution functions have been tested to determine which ones best model daily, running 7-day average and running 30-day average effluent BOD_5 and SS concentration data. The time series nature of daily effluent data has been investigated by testing the data for randomness and by testing for the existence of annual and weekly cycles. Effects of influent, operational, and environmental parameters have been examined by using multiple linear regression analysis to correlate linear combinations of various process parameters with effluent BOD_5 and SS concentrations. Finally, the percentage of the variability in effluent concentration data caused by measurement errors was estimated from information on the precision of the BOD_5 and SS tests.

EFFLUENT DATA DISTRIBUTIONS

The characterization of daily, running 7-day average and running 30-day average effluent BOD_5 and SS concentration data was conducted in three steps. First, summary statistics for each type of effluent data set were computed for each treatment plant. Second, coefficients for two empirical distribution models were computed. Third, two empirical and three theoretical distribution models were tested to determine their usefulness in modeling effluent data from each plant.

Summary Statistics

A useful first step in characterizing the distribution of a set of data is to examine summary statistics such as the mean, standard deviation and skew coefficient. Summary statistics for trickling filter effluent data are presented in Tables 6a, 6b, and 6c.

Some general characteristics of trickling filter effluent dta can be seen from these results. For the daily data, the averages of mean BOD_5 a from all plants and the mean SS concentration from all plants are approximately equal. However, the average standard deviation for SS data is somewhat higher than the average standard deviation for BOD_5 data. Thus it may be concluded that, overall, BOD_5 and suspended solids data tend to be of the same average magnitude but that SS data tend to be somewhat more variable.

EPA Form 2350-4 (4-80)
(PREVIOUSLY CIN. EPA FORM 287)
TYPING GUIDE SHEET

TABLE 6A

SUMMARY STATISTICS FOR DAILY EFFLUENT BOD_5 AND SUSPENDED SOLIDS CONCENTRATIONS

	BOD_5							Suspended Solids						
Plant Number	Mean g/m^3	Sx g/m^3	V_x	Skew	Max g/m^3	Min g/m^3	Valid Obs	Mean g/m^3	Sx g/m^3	V_x	Skew	Max g/m^3	Min g/m^3	Valid Obs
1	33.31	15.28	0.46	0.92	102.0	8.0	357	52.50	31.17	0.59	2.26	280.0	6.0	356
2	10.73	7.24	0.67	1.62	43.0	0.0	365	21.45	13.35	0.62	1.15	80.0	1.0	366
3	10.05	7.52	0.75	1.68	61.0	1.0	362	21.04	16.69	0.79	1.17	88.0	1.0	361
4	58.35	20.83	0.36	0.10	107.0	22.0	250	54.85	15.02	0.27	0.10	94.0	20.0	250
5	29.24	11.21	0.38	1.31	93.0	6.0	492	18.34	8.81	0.48	2.16	84.0	3.0	493
6	27.04	6.75	0.25	0.09	58.0	3.0	365	15.09	5.98	0.40	0.40	37.0	2.0	365
7	23.21	16.93	0.73	1.62	100.0	1.0	383	24.05	15.09	0.63	0.99	86.0	2.0	386
8	43.09	6.59	0.15	0.61	69.0	28.0	365	34.03	13.63	0.40	4.56	172.0	12.0	365
9	51.13	19.95	0.39	0.23	125.0	12.0	359	41.08	17.61	0.43	1.60	130.0	12.0	365
10	21.02	6.22	0.30	-0.15	33.0	8.0	250	23.60	14.42	0.61	1.08	86.0	2.0	250
11	18.31	11.98	0.65	2.17	104.0	0.0	355	16.15	8.85	0.55	1.09	64.0	1.0	363
Average	29.59	11.86	0.46	0.93	--	--	--	29.29	14.60	0.52	1.51	--	--	--
Median	--	11.21	0.39	0.92	93.0	6.0	--	--	14.42	0.55	1.15	86.0	2.0	--

TABLE 6B

SUMMARY STATISTICS FOR CONTINUOUS 7-DAY AVERAGE EFFLUENT BOD AND SUSPENDED SOLIDS CONCENTRATIONS

	BOD_5							Suspended Solids						
Plant Number	Mean g/m^3	S_x g/m^3	V_x	Skew	Max g/m^3	Min g/m^3	Valid Obs	Mean g/m^3	S_x g/m^3	V_x	Skew	Max g/m^3	Min g/m^3	Valid Obs
1	33.12	11.52	0.35	0.15	62.3	14.0	360	52.52	18.63	0.35	0.38	103.6	17.6	360
2	10.56	4.95	0.47	0.93	25.4	3.1	360	21.22	9.18	0.43	0.70	49.1	6.1	360
3	9.99	4.76	0.48	0.67	24.6	1.7	360	21.14	8.00	0.38	0.26	41.5	5.1	360
4	58.31	20.17	0.35	0.07	96.8	27.2	360	54.82	12.65	0.23	-0.03	78.8	30.0	360
5	29.15	6.86	0.24	0.57	53.1	14.0	514	18.41	5.26	0.29	0.73	39.0	7.3	516
6	27.02	3.86	0.14	-0.39	35.7	14.4	359	15.07	4.40	0.29	0.04	27.4	6.4	359
7	22.87	13.32	0.58	1.48	81.7	7.1	391	23.62	9.84	0.42	0.59	52.3	5.9	391
8	43.00	4.79	0.11	0.45	54.3	34.1	360	34.04	7.17	0.21	1.61	63.1	24.1	360
9	51.49	17.47	0.34	-0.07	82.9	18.3	359	41.23	12.57	0.30	0.73	80.3	18.3	359
10	21.01	4.59	0.22	0.21	29.6	10.0	359	23.86	9.59	0.40	1.21	59.2	6.4	359
11	17.99	8.40	0.47	1.28	48.7	4.5	358	16.11	5.07	0.31	0.60	34.7	5.9	359
Average	—	9.15	0.34	0.49	—	—	—	—	9.31	0.33	0.62	—	—	—
Median	—	6.86	0.35	0.45	53.1	14.0	—	—	9.18	0.31	0.60	52.3	6.4	—

TABLE 6C

SUMMARY STATISTICS FOR RUNNING 30-DAY AVERAGE EFFLUENT BOD AND SUSPENDED SOLIDS CONCENTRATIONS

	BOD_5							Suspended Solids						
Plant Number	Mean g/m^3	Sx g/m^3	V_x	Skew	Max g/m^3	Min g/m^3	Valid Obs	Mean g/m^3	Sx g/m^3	V_x	Skew	Max g/m^3	Min g/m^3	Valid Obs
1	32.11	9.91	0.31	0.15	51.9	17.0	337	51.59	15.63	0.30	0.78	88.9	29.7	337
2	10.00	3.53	0.35	1.04	20.7	5.4	337	20.93	7.66	0.37	0.93	41.1	9.5	337
3	9.77	3.69	0.38	0.77	19.6	4.2	337	21.05	5.47	0.26	0.09	31.5	11.0	337
4	57.69	19.85	0.34	0.11	88.9	32.0	337	54.30	11.54	0.21	-0.02	72.2	37.3	337
5	29.03	5.16	0.18	0.57	42.7	20.0	516	18.58	3.81	0.21	0.27	30.5	11.2	516
6	26.78	2.39	0.09	-0.74	31.0	20.2	336	14.85	3.88	0.26	-0.05	21.1	7.47	336
7	22.59	11.63	0.51	1.09	52.1	10.1	368	23.22	7.55	0.33	0.16	38.6	11.2	368
8	42.75	3.98	0.09	0.57	51.7	36.8	337	34.01	4.50	0.13	-0.07	43.1	26.0	337
9	52.94	15.56	0.29	-0.16	79.0	25.1	336	41.81	9.79	0.23	0.40	61.7	26.8	336
10	21.10	3.48	0.16	0.66	27.9	15.7	336	23.99	6.25	0.26	0.23	39.4	11.3	336
11	17.05	6.18	0.36	0.84	34.5	7.8	336	15.81	3.44	0.22	-0.06	24.9	7.9	336
Average	—	7.76	0.28	0.45	—	—	—	—	7.23	0.25	0.24	—	—	—
Median	—	5.16	0.31	0.57	42.7	17.0	—	—	6.25	0.26	0.16	39.4	11.2	—

It can be seen in Table 6a that the skew coefficient is positive for every set of daily effluent concentration data except one. The distributions of daily effluent BOD_5 and SS concentrations generally are not symmetrical, but have larger ranges of values above than below the distribution mode.

Empirical Distribution Models

Two types of empirical distribution functions were developed to model effluent concentration data. The first is a one-parameter model in which a series of linear regression equations is used to predict percentile values based on the mean value of the data series. The second type is like the one-parameter model, except that it is a two-parameter model and uses the mean and standard deviation to predict percentile values. In Table 7, regression coefficients, coefficients of determination, and standard errors are presented for the regression equations of both models for daily effluent BOD_5. Coefficients for the daily SS model, as well as the models for running 7-day and 30-day averages are presented in Appendix A.

The results for the daily effluent BOD_5 model are representative of the results for the other models. As can be seen from the values of r^2 and the standard error, these models, especially the two-parameters models, provide a fairly accurate description of the distribution of effluent BOD_5 and SS data. For the purpose of predicting percentile values, these models appear most accurate between the 10th and 95th percentiles and become less accurate for very high or very low percentiles.

If the mean value for effluent BOD_5 or SS concentrations were small enough, these empirical models would predict negative concentrations for some percentiles, which is, of course, a physical impossibility. The possibility of negative values emphasizes the fact that these models are only valid over the range of data used to compute the model coefficients.

Comparison of Distibution Models

Five probability distributions have been tested for their fit to effluent BOD_5 and SS data using the Kolmogorov-Smirnov (K-S) goodness-of-fit test. The following distributions were tested: the one and two-parameter empirical models discussed above and the noraml, log-normal and gamma distribution functions. The results of these tests are presented in Tables 8a and 8f and in Table 9.

In Tables 8a through 8f, the distributions which are not rejected at the 1 percent significance level are indicated for each plant. A 1 percent significance level has been used as a rejection criteria because of the large sample sizes (greater than 240). As sample size increases, the probability decreases that a data set will pass a goodness-of-fit test for a hypothesized distribution. (5) The objective in this study is not to prove that the shapes of hypothesized and observed distributions are exactly the same, but to compare the usefulness of various models for describing effluent data distributions. It has been shown (for activated sludge plant effluent data) that with a sample size of approximately 350, a hypothesized distribution that cannot be rejected at the 1 percent level is adequate for describing the observed data set. (33)

The average values of the K-S test statistic, D, are shown in Table 9. For each test of the fit of a data set and a hypothesized distribution, D is equal to

BEGIN FIRST LINE OF TEXT HERE
CENTER OF PAGE
TOP OF IMAGE AREA
DROPPED HEAD BEGIN SECTIONS HERE

TABLE 7

COEFFICIENTS AND STATISTICAL RESULTS FOR ONE AND TWO-PARAMETER EMPIRICAL DISTRIBUTION MODELS, DAILY EFFLUENT BOD$_5$ CONCENTRATIONS

	One-Parameter Model				Two-Parameter Model				
Percentile	A_1	C_1 g/m^3	R^2	S_y g/m^3	A_2	B_2	C_2 g/m^3	R^2	S_y g/m^3
0	0.51	-6.3	0.66	5.7	0.72	-1.00	-1.3	0.86	3.9
10	0.66	-2.9	0.82	4.9	0.90	-1.09	2.6	0.99	1.4
20	0.74	-2.3	0.90	3.9	0.93	-0.86	2.1	0.99	1.0
30	0.81	-1.8	0.93	3.6	0.98	-0.78	2.1	0.99	1.2
40	0.91	-2.0	0.97	2.5	1.03	-0.55	0.8	0.99	0.9
50	1.03	-2.4	0.99	1.7	1.09	-0.26	-1.1	0.99	1.3
60	1.16	-2.8	0.99	2.2	1.12	0.19	-3.8	0.99	2.2
70	1.23	-1.5	0.98	2.6	1.13	0.42	-3.6	0.99	2.1
80	1.29	0.2	0.97	3.5	1.13	0.74	-3.6	0.99	1.5
90	1.29	6.4	0.92	5.9	0.99	1.34	-0.4	0.99	1.1
95	1.27	10.9	0.86	8.0	0.88	1.78	1.9	0.99	2.4
98	1.23	16.8	0.77	10.8	0.72	2.32	5.0	0.97	4.1
99	1.19	25.9	0.60	15.7	0.51	3.10	10.2	0.89	8.5
100	1.05	48.2	0.31	25.1	0.00	4.76	24.1	0.78	15.0

---

[a] Coefficients for other data analyses are found in Appendix A.

BEGIN LAST LINE OF TEXT
PAGE NUMBER
EPA Form 2350-4 (4-00)
TYPING GUIDE SHEET

TABLE 8A

TYPE I ERROR RESULTS OF KOLMOGOROV-SMIRNOV TEST; DAILY EFFLUENT BOD$_5$ CONCENTRATIONS

Plant Number	2-Tailed P					Accepted Distributions[a]				
	Normal	Log-Normal	Gamma	One Parameter Empirical	Two Parameter Empirical	Normal	Log-Normal	Gamma	One Parameter Empirical	Two Parameter Empirical
1	0.007	0.118	0.562	0.003	0.569		x	x		x
2	0	0.085	0.008	0.004	0.001		x			
3	0	0.005	0.055	0	0			x		
4	0.014	0.001	0.022	0.009	0.187	x		x		x
5	0.044	0.041	0.391	0.216	0.338	x	x	x	x	x
6	0.192	0.001	0.050	0	0.010	x		x		x
7	0	0.735	0.110	0	0.023		x	x		x
8	0.162	0.449	0.437	0	0.000	x	x	x		
9	0.014	0.001	0.002	0.002	0.365	x				x
10	0.004	0	0	0	0					
11	0	0.002	0.007	0.001	0.007					

[a] Indicated distributions not rejected at 1% level of significance.

TABLE 8B

TYPE I ERROR RESULTS OF KOLMOGOROV-SMIRNOV TEST; DAILY EFFLUENT SUSPENDED SOLIDS CONCENTRATIONS

	2-Tailed P					Accepted Distributions[a]				
Plant Number	Normal	Log-Normal	Gamma	One Parameter Empirical	Two Parameter Empirical	Normal	Log-Normal	Gamma	One Parameter Empirical	Two Parameter Empirical
1	0	0.027	0.018	0	0.674		x	x		x
2	0	0.086	0.660	0.005	0.183		x	x		x
3	0	0.001	0.014	0	0.011			x		x
4	0.333	0.015	0.067	0.001	0.012	x	x	x		x
5	0	0	0.003	0	0.003					
6	0.200	0	0.032	0	0.097	x		x		x
7	0.037	0.001	0.109	0	0.083	x		x		x
8	0	0.057	0	0	0		x			
9	0	0.583	0.148	0.361	0.162		x	x	x	x
10	0.076	0.013	0.601	0.013	0.460	x	x	x	x	x
11	0.009	0.011	0.403	0.038	0.701		x	x	x	x

[a] Indicated distributions not rejected at 1% significance level.

TABLE 8C

TYPE I ERROR RESULTS OF KOLMOGOROV-SMIRNOV TEST; 7-DAY AVERAGE EFFLUENT BOD$_5$ CONCENTRATIONS

	2-Tailed P					Accepted Distributions[a]				
Plant Number	Normal	Log-Normal	Gamma	One Parameter Empirical	Two Parameter Empirical	Normal	Log-Normal	Gamma	One Parameter Empirical	Two Parameter Empirical
1	0.006	0	0	0	0.015					x
2	0	0.164	0.054	0	0		x	x		
3	0	0.221	0.098	0	0.002		x	x		
4	0	0	0	0	0.046					x
5	0.041	0.233	0.360	0	0.606	x	x	x		x
6	0.219	0.009	0.050	0	0.004	x		x		
7	0	0.079	0.029	0	0		x	x		
8	0.041	0.230	0.162	0	0	x	x	x		
9	0.001	0	0	0	0.139					x
10	0.027	0.049	0.038	0	0.042	x	x	x		x
11	0	0.250	0.019	0	0.040		x	x		x

[a] Indicated distributions not rejected at 1% significance level.

TABLE 8D

TYPE I ERROR RESULTS OF KOLMOGOROV-SMIRNOV TEST; 7-DAY AVERAGE EFFLUENT SUSPENDED SOLIDS CONCENTRATIONS

	2-Tailed P					Accepted Distributions[a]				
Plant Number	Normal	Log-Normal	Gamma	One Parameter Empirical	Two Parameter Empirical	Normal	Log-Normal	Gamma	One Parameter Empirical	Two Parameter Empirical
1	0	0.095	0.023	0.003	0.469		x	x		x
2	0.007	0.652	0.293	0	0.216		x	x		x
3	0.172	0.184	0.193	0.155	0.465	x	x	x	x	x
4	0.002	0	0	0.002	0.010					x
5	0.029	0.067	0.349	0.002	0.183	x	x	x		x
6	0.022	0.006	0.011	0.003	0.002	x		x		
7	0.009	0.391	0.709	0.001	0.146		x	x		x
8	0.002	0.064	0.016	0	0		x	x		
9	0.059	0.845	0.904	0.722	0.551	x	x	x	x	x
10	0.011	0.319	0.146	0.231	0.081	x	x	x	x	x
11	0.594	0.097	0.374	0.018	0.605	x	x	x	x	x

[a] Indicated distributions not rejected at 1% significance level.

TABLE 8E

TYPE I ERROR RESULTS OF KOLMOGOROV-SMIRNOV TEST; 30-DAY AVERAGE EFFLUENT BOD_5 CONCENTRATIONS

	2-Tailed P					Accepted Distributions[a]				
Plant Number	Normal	Log-Normal	Gamma	One Parameter Empirical	Two Parameter Empirical	Normal	Log-Normal	Gamma	One Parameter Empirical	Two Parameter Empirical
1	0	0	0	0	0.002					
2	0.001	0.003	0.012	0	0			x		
3	0	0.056	0.010	0	0		x	x		
4	0	0	0	0	0.001					
5	0.053	0.125	0.151	0	0.062	x	x	x		x
6	0.001	0	0	0	0					
7	0	0	0	0	0					
8	0.020	0.046	0.059	0	0	x	x	x		
9	0.011	0	0	0.001	0	x				
10	0	0.001	0	0	0.006					
11	0.044	0.191	0.116	0	0.131	x	x	x		x

[a] Indicated distributions not rejected at 1% significance level.

TABLE 8F

TYPE I ERROR RESULTS OF KOLMOGOROV-SMIRNOV TEST; 30-DAY AVERAGE EFFLUENT SUSPENDED SOLIDS CONCENTRATIONS

	2-Tailed P					Accepted Distribution[a]				
Plant Number	Normal	Log-Normal	Gamma	One Parameter Empirical	Two Parameter Empirical	Normal	Log-Normal	Gamma	One Parameter Empirical	Two Parameter Empirical
1	0	0	0	0.001	0.044					x
2	0	0.167	0.024	0	0.044		x	x		x
3	0.016	0.041	0.016	0.105	0.167	x	x	x	x	x
4	0	0	0	0	0					
5	0.148	0.165	0.429	0.002	0	x	x	x		
6	0.002	0.018	0.003	0.013	0.008		x		x	
7	0	0.003	0.003	0	0.005					
8	0.001	0	0	0	0					
9	0.006	0.059	0.043	0.032	0.024		x	x	x	x
10	0.074	0.001	0.003	0.088	0.056	x			x	x
11	0.001	0	0	0	0.142					x

[a] Indicated distributions not rejected at 1% significance level.

BEGIN FIRST LINE OF TEXT HERE
CENTER OF PAGE
TOP OF IMAGE AREA
DROPPED HEAD: BEGIN SECTIONS HERE

TABLE 9

AVERAGE MAXIMUM PERCENTILE DIFFERENCES[a]

Effluent Parameter	Distributions				
	Normal	Log-Normal	Gamma	One Parameter Empirical	Two Parameter Empirical
Daily BOD_5	10.1	8.3	7.7	13.9	8.9
Daily SS	10.3	8.7	7.7	12.4	7.3
7-Day Avg. BOD_5	9.8	8.1	8.3	17.7	9.0
7-Day Avg. SS	7.7	6.3	6.2	9.2	6.5
30-Day Avg. BOD_5	11.8	10.3	10.5	22.3	13.7
30-Day Avg. SS	10.7	10.4	10.4	11.0	9.8

[a] Each entry is the average for 11 treatment plants of the K-S test statistic, D. The statistic D is equal to the maximum percentile difference between the hypothesized distribution and the distribution of the data.

BEGIN LAST LINE OF TEXT

the maximum percentile difference between the observed and hypothesized distribution. For each effluent parameter, the average value of D for the 11 plants studied has been computed for each distribution.

It can be seen from the results in Tables 8a through 8f that for any single effluent parameter, (e.g. daily BOD$_5$, 7-day average suspended solids, etc.) the distributions that best fit the data change from treatment plant to treatment plant. However, it can be concluded that, in general, the log-normal, gamma and two-parameter empirical distribution functions are equivalent for the purpose of describing effluent data distributions. This conclusion is based on the number of accepted distributions at 1 percent significance level.

The individual calculations for each Kolmogorov-Smirnov test were examined and it was found that the greatest differences between the predicted and observed precentiles generally occured between the 30th and 70th percentiles, and that the percentile differences between the tails of the hypothesized and observed distributions were always much smaller than the maximum difference. Thus, if one of the distributions is used to estimate the frequency of occurence of values equal to or greater than a data value above the 80th percentile, the percentile error would be smaller than the value of test statistic, D.

RANDOMNESS OF EFFLUENT DATA

The hypothesis that . . . trickling filter effluent BOD$_5$ and SS concentration data consist of series of purely random numbers has been tested using a runs test and a test for serial correlation. In addition to simply determining if the data are random, these tests can be used to determine the relative magnitude of non-randomness in each set of data and provide a basis for comparing the relative randomness of different sets of data.

The results of the randomness tests of daily effluent BOD and SS data are presented in Table 10. Using a 5 percent significance level as the rejection criterion, the hypothesis of randomness can be rejected for every set of BOD and SS data on the basis of at least one of the two tests. From the magnitude of the correlation coefficients and Z statistics, it can be concluded that the BOD data from every plant and the SS data from every plant but one are highly non-random. The Z statistic for the runs test is negative for every data set, which indicates that all of the data have fewer runs than would be expected for random data.

The SS data from Plant 3 can be considered approximately random. For this data set, the hypothesis of randomness cannot be rejected at the 5 percent significance level on the basis of the runs test. Furthermore, the serial correlation coefficient, though statistically significant, is small enough to be considered of little practical significance.

For nine out of the eleven treatment plants, the Z statistic and the serial correlation coefficient are closer to zero for SS data than for BOD$_5$ data. Thus it can be concluded that, for most of the treatment plants studied, effluent BOD$_5$ data are less random than SS data.

TABLE 10

RESULTS OF RANDOMNESS TESTS FOR DAILY EFFLUENT BOD$_5$ AND SUSPENDED SOLIDS

	BOD$_5$				Suspended Solids			
	Runs Test		Serial Correlation		Runs Test		Serial Correlation	
Plant Number	Z	Significance level for 2-Tailed P, percent	rρ	Significance level for 1-Tailed P percent	Z	Significance level for 2-Tailed P, percent	rρ	Significance level for 1-Tailed P, percent
1	-10.5	0.0	0.65	<0.1	-4.6	0.0	0.31	0.1
2	-10.2	0.0	0.57	<0.1	-9.1	0.0	0.57	0.1
3	-5.9	0.0	0.39	<0.1	-1.3	18.8	0.12	2.1
4	-13.8	0.0	0.95	<0.1	-10.0	0.0	0.71	0.1
5	-7.1	0.0	0.39	<0.1	-6.5	0.0	0.34	0.1
6	-6.1	0.0	0.32	<0.1	-10.0	0.0	0.54	0.1
7	-8.9	0.0	0.56	<0.1	-6.4	0.0	0.47	0.1
8	-7.2	0.0	0.50	<0.1	-6.2	0.0	0.36	0.1
9	-15.2	0.0	0.84	<0.1	-9.5	0.0	-0.62	0.1
10	-6.1	0.0	0.53	<0.1	-2.8	0.5	0.26	0.1
11	-11.0	0.0	0.62	<0.1	-6.0	0.0	0.35	0.1

ANNUAL AND WEEKLY CYCLES

It may be concluded from the results of the randomness tests that there are significant non-random patterns in trickling filter effluent data. The existence of annual or weekly cycles in the data has been examined by performing a two-way analysis of variance on BOD_5 and suspended solids data from each plant and by pooling normalized data from all the treatment plants and plotting the averages for each month and day of the week.

Analaysis of Variance

Two-way analysis of variance was used to test the following hypotheses for effluent BOD_5 and SS concentration data from each treatment plant:

(1) Effluent BOD_5 (SS) concentrations are not dependent on the month in which they occur.

(2) Effluent BOD_5 (SS) concentrations are not dependent on the day of the week on which they occur.

(3) If effluent BOD_5 (SS) concentrations are dependent on month and day of the week, the effect of month and the effect of day of the week are independent and additive.

A 5 percent significance level is used for most of the statistical tests in this study. However, as was explained in the Statistical Methods section, a stricter 1 percent significance level is used for hypotheses (1) and (2) above to adjust for the fact that the data used in this study do not strictly conform to the assumptions upon which analysis of variance is based. Because rejection of hypothesis (3) is an unfavorable result, a less strict 5 percent level will be used; to be conservative.

The results of the two-way analysis of variance calculations are presented in Tables 11a and 11b. Also given in these tables are the ratios of the maximum group average to the minimum group average for data grouped by month and for data grouped by day of the week.

The first hypothesis (no effect of month) can be rejected at the 1 percent significance level for effluent BOD_5 and SS data from every treatment plant (Tables 11a and 11b). In other works, for each treatment plant studied, effluent BOD_5 and suspended solids concentrations are in some manner dependent on the month in which they occur. As can be seen from the ratios of maximum monthly average to minimum monthly average, the differences between monthly averages are of physical as well as of statistical significance.

Because only one year of data from each plant was used, it may not be concluded that effluent concentrations at these 11 treatment plants have repeating annual cycles. There is no way of telling from the available data whether particular sequences of monthly averages are repeated yearly or are merely random fluctuations in average monthly performance. The only conclusion that can be made is that the average monthly effluent BOD_5 and SS concentrations do change significantly over the year studied.

TABLE 11A

RESULTS OF 2-WAY ANALYSIS OF VARIANCE FOR EFFLUENT BOD_5

	Effect of Month		Effect of Day		Interaction Effect
Plant Number	Probability that data are random	Ratio $\frac{Max}{Min}$	Probability that data are random	Ratio $\frac{Max}{Min}$	Probability that data are random
1	0.0	3.0	0.0	1.5	9.5
2	0.0	3.5	76.3	1.2	48.5
3	0.0	3.4	1.4	1.5	37.4
4	0.0	2.6	—	—	—
5	0.0	1.6	23.7	1.1	80.5
6	0.0	1.4	0.3	1.2	91.6
7	0.0	4.8	53.1	1.2	87.5
8	0.0	1.3	0.0	1.2	14.2
9	0.0	2.9	73.7	1.1	98.5
10	0.0	1.7	—	—	—
11	0.0	3.5	0.0	2.0	0.7

TABLE 11B

RESULTS OF 2-WAY ANALYSIS OF VARIANCE FOR EFFLUENT SOLIDS

	Effect of Month		Effect of Day		Interaction Effect
Plant Number	Probability that data are random	Ratio $\frac{\text{Max}}{\text{Min}}$	Probability that data are random	Ratio $\frac{\text{Max}}{\text{Min}}$	Probability that data are random
1	0.0	2.2	0.0	1.6	4.7
2	0.0	3.9	74.8	1.1	51.9
3	0.0	2.4	0.0	1.9	95.7
4	0.0	1.9	—	—	—
5	0.0	1.9	8.7	1.1	31.5
6	0.0	2.5	65.5	1.1	99.8
7	0.0	3.0	15.3	1.3	90.0
8	0.0	1.4	0.0	1.5	94.5
9	0.0	2.0	18.7	1.2	86.6
10	0.0	2.7	—	—	—
11	0.0	2.1	0.0	1.6	23.5

The second hypothesis (no effect of day of the week) can be rejected at the 1 percent significance level for BOD_5 data from four treatment plants and for SS data from four treatment plants. (Tables 11a and 11b). In othe words, the effluent concentrations in these eight sets of data are dependent on the day of the week on which they occur. Because there are 52 weeks in each set of data, it may be concluded that there is a repeating weekly cycle in each of these eight data sets.

The third hypothesis (no interaction effect) can be rejected at the 5 percent level for one set of BOD_5 data and for one set of SS data, which indicates that the effects of month and day of the week are not independent of one another for these two sets of data. One possible cause for these interaction effects could be that the amplitude of the weekly cycles in these data sets changes from month to month. Another possible cause could be that the shape of the weekly cycle changes from month to month. However, it is beyond the scope of this study to examine each set of data to determine the precise pattern of weekly cycles and the reasons for any significant interaction effects.

Plots of Pooled Data

The existence of any general weekly or annual patterns in the effluent data from all the treatment plants was studied by plotting monthly and daily averages for pooled data. Effluent BOD_5 and SS concentration data from each treatment plant were first normalized. Then, the normalized data from all of the treatment plants were pooled together and grouped according to the month in which the data were taken. Plots of the monthly averages for effluent BOD_5 and SS are presented in Figure 5. To study weekly patterns, effluent data from the nine treatment plants with daily data were pooled and then grouped according to the day of the week on which the data occured. Plots of the averages for each day of the week for effluent BOD_5 and SS are presented in Figure 6.

As can be seen in Figure 5, effluent BOD_5 effluent BOD_5 and SS data tend to have annual cycles that are roughly in phase with one another. Concentrations of both BOD_5 and SS tend to be higher than average in the winter months and lower than average in the summer and early fall.

It should be emphasized that, for individual treatment plants, there are considerable deviations from the patterns of monthly averages shown in Figure 5. To illustrate this point, the deviations of each monthly average from the annual mean are plotted in Appendix B for BOD_5 data from each of the 11 treatment plants studied. Although there is considerable scatter in the data, the tendency for above average concentrations in the winter and below average concentrations in the summer and early fall can still be seen.

The amplitudes of the weekly patterns for pooled effluent data (Figure 6) are much smaller than the amplitudes of the monthly patterns of pooled data. There appears to be tendency for effluent BOD_5 and SS concentrations to be below average on Sundays and for effluent suspended solids concentrations to be above average on Wednesdays.

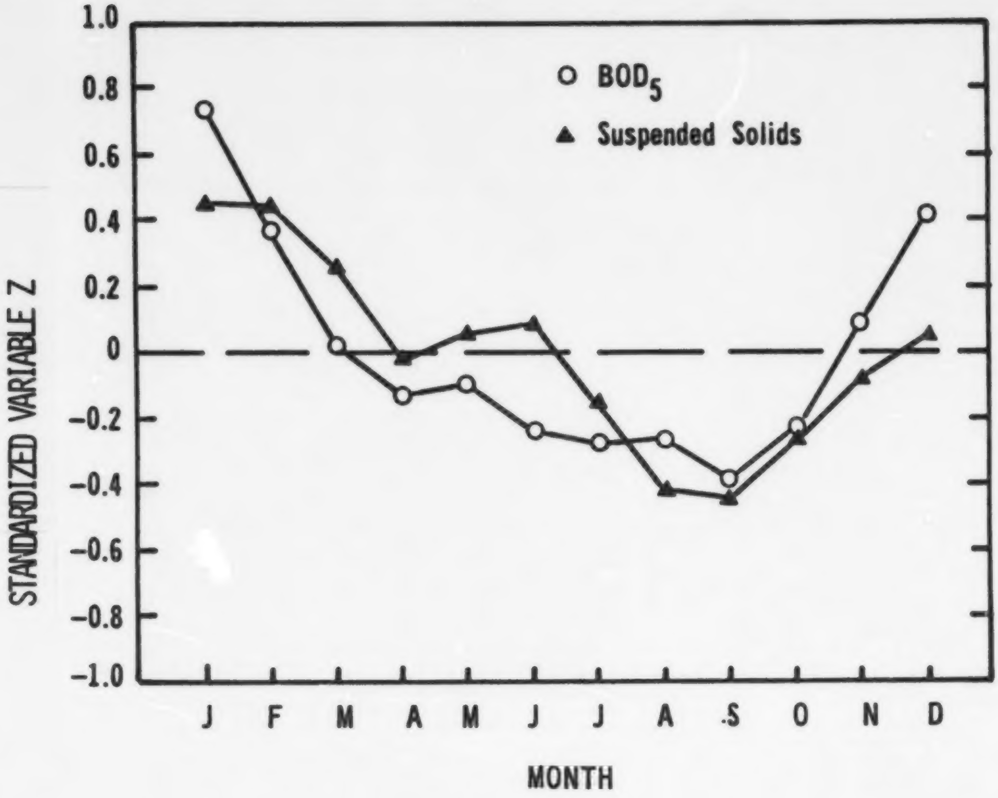


Figure 5. **Annual pattern of normalized effluent data pooled from 11 plants.**

BEGIN FIRST LINE OF TEXT HERE
CENTER OF PAGE
TOP OF IMAGE AREA
DROPPED HEAD, BEGIN SECTIONS HERE

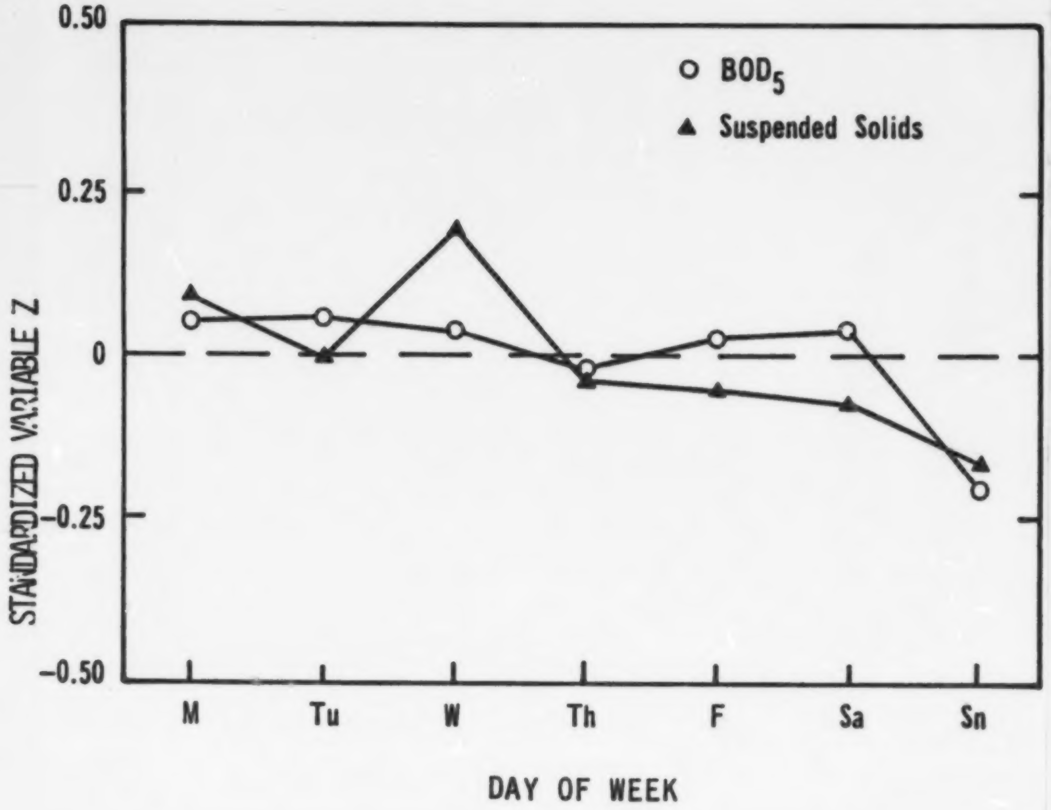


Figure 6. Weekly pattern of normalized effluent data pooled from 9 plants.

BEGIN LAST LINE OF TEXT
BOTTOM OF IMAGE AREA OUTSIDE DIMENSION FOR TABLES AND ILLUSTRATIONS

PAGE NUMBER
EPA Form 2350-4 (4-80)
(PREVIOUSLY CIN: EPA FORM 227)
U.S. GOVERNMENT PRINTING OFFICE 1980
TYPING GUIDE SHEET

TABLE 12A

COEFFICIENTS FOR PART 1 REGRESSION MODELS PREDICTING EFFLUENT BOD_5 CONCENTRATIONS[a]

Plant Number	Q	BOD_i	SS_i	BOD_p	SS_p	$WTEMP^{-1}$	$ATEMP^{-1}$	R	pH-7.0	Non Normal-lized Sy g/m^3	R^2
1	-0.27_5	0.47_1	$*-0.15_6$			0	0.16_3	-0.40_2	$*+0.21_4$	13.1	0.27
2	0	0.16_3	0	0.45_1	0	0.24_2				5.9	0.34
3	0	0	0	0.36_1	0	0.32_2	0		0	6.3	0.27
4	0.09_3	-0.07_4	0	0.72_1	0.04_5	0.26_2				6.4	0.91
5	0.17_3	0.18_2	0	0.43_1	0	0		0.17_4	0	8.9	0.39
6	0	0	0	0.27	0	0				6.5	0.07
7	-0.17_4	0.20_1	0	0.30_3	$*-0.10_5$	0.29_2	0.09_6		0	13.1	0.43
8	0.21_3	0.24_2	0			0.30_1				6.1	0.15
9	0	0.17_3	$*-0.13_4$	0.64_1	0.19_2					12.1	0.64
10	0.28_2	0	0	0.36_1	0	-0.22_3		0		5.6	0.22
11	$*-0.15_3$	$*-0.11_2$	$*-0.11_5$	0.52_1	0.17_4					10.7	0.23

[a] Subscripts indicate order in which variables were entered into the regression equation. Variables for which data was available but which were not entered into regression equation due to stepwise criteria are indicated by "0". A blank entry indicates data for that variable was not available. Limiting F value for stepwise criteria was 2.6. A "*" indicates that the sign of the simple correlation coefficient between BOD_5 and the indicated variable is the opposite of the sign of the multiple regression coefficient.

TABLE 12B

COEFFICIENTS FOR PART 1 REGRESSION MODELS PREDICTING EFFLUENT SUSPENDED SOLIDS CONCENTRATION[a]

Plant Number	Q	BOD_i	SS_i	BOD_p	SS_p	$WTEMP^{-1}$	$ATEMP^{-1}$	R	pH-7.0	Non Normal-lized S_y g/m^3	R^2
1	$*-0.20_5$	0	0.23_3			0	0.14_4	-0.31_1	0.30_2	28.6	0.17
2	0.29_3	0	0.11_4	0.27_2	0.26_1	$*-0.12_5$				12.2	0.18
3	0.15_3	0	0.25_2	0	0.25_1	0.09_4	0		0	15.1	0.19
4	0.13_3	0	0	0.30_2	0.47_1	0.15_4				8.5	0.69
5	0	0	0.13_4	0	0.27_1	0.22_2		0.13_3	0	8.0	0.18
6	0	$*-0.10_4$	0.19_3	0	0.36_2	0.53_1				5.0	0.30
7	-0.14_3	0	0.11_4	0	0.16_2	0.28_1	0		0	13.9	0.17
8	0.16_3	-0.37_2	0.57_1			0.10_4				10.4	0.42
9	0	0	0	0	0.68_1					13.4	0.47
10	0.23_3	0	0	0	0.33_2	0.23_1		0		13.1	0.18
11	0	$*-0.20_2$	0	0	0.65_1					7.2	0.34

[a] Subscripts indicate order in which variables were entered into the regression equation. Variable for which data was available but which were not entered into regression equation due to stepwise criteria are indicated by "0". A blank entry indicates data for that variable was not available. Limiting F value for stepwise criteria was 2.6. A "*" indicates that the sign of the simple correlation coefficient between SS_e and the indicated variable is the opposite of the sign of the multiple coefficient.

TABLE 13A

RESULTS OF PART II REGRESSION ANALYSIS, EFFLUENT BOD_5

Plant Number	Variable in the Order Entered into Regression Equation[a]						Non Normal-lized Sy g/m^3	R^2
	1	2	3	4	5	6		
1	AVTEMP (-)	$QBOD_i$	$ATEMP^2$ (-)	SS_i			11.3	0.46
2	BOD_p^2	WTEMP (-)	BOD_i, t-1	SS_i, t-2	BOD_i	Q (-)	5.8	0.37
3	BOD_p^2	$WTEMP^{-1}$	SS_i t-1 (-)	DTEMP			6.2	0.27
4	BOD_p	WTEMP (-)	$QBOD_p$	BOD_i	$QBOD_i$	Q^2	6.2	0.91
5	BOD_p	BOD_p, t-1	BOD_p^2	$QBOD_i$	R^2	BOD_i, t-1	8.5	0.46
6	BOD_p	BOD_p, t-2					6.5	0.09
7	DTEMP	$QBOD_p$	BOD_p, t-2	BOD_i^2	$ATEMP_{t-1}$ (-)	SS_p	12.6	0.48
8	WTEMP (-)	$QBOD_i$	SS_i, t-1 (-)	SS_i^2			5.7	0.25
9	BOD_p	BOD_p, t-2	Q_{t-2}	SS_p, t-2	$QBOD_i$	SS_i^2	11.6	0.67
10	$QBOD_p$	$WTEMP^2$					5.5	0.24
11	BOD_p	BOD_p, t-1	Q_{t-2} (-)	BOD_i, t-1	$QBOD_i$		9.0	0.35

---

[a] A minus sign in parentheses, "(-)", indicates that the sign of the simple correlation coefficient between the preceding variable and BOD_5 is negative; if no sign is shown, the correlation is positive.

TABLE 13B

RESULTS OF PART II REGRESSION ANALYSIS FOR EFFLUENT SUSPENDED SOLIDS

Plant Number	Variable in the Order Entered into Regression Equation[a]						Non Normalized Sy g/m^3	R^2
	1	2	3	4	5	6		
1	WTEMP (-)	QSS_i					27.2	0.26
2	$QBOD_p$	SS_p	QSS_i	pH			11.8	0.24
3	QSS_p	SS_i	SS_p^2	SS_p, t-1	SS_p, t-2	SS_i^2	14.6	0.27
4	SS_p	SS_p^2	BOD_p	Q^2	$WTEMP^2$ (-)		7.7	0.74
5	SS_p	$WTEMP^2$ (-)	SS_i, t-1	SS_i	R^2	SS_p^2	6.5	0.30
6	WTEMP (-)	SS_p	SS_p^2	Q_{t-2} (-)	SS_p, t-2	SS_i	4.8	0.37
7	Q_{t-1} (-)	WTEMP (-)	BOD_p, t-2	QSS_i	SS_i^2	SS_p	13.5	0.23
8	QSS_i	$QBOD_i$ (-)	WTEMP (-)	SS_i	Q^2	SS^2	9.9	0.49
9	SS_p	SS_p, t-1	BOD_i, t-2				13.2	0.49
10	QSS_p	WTEMP (-)					13.2	0.17
11	SS_p	BOD_i	SS_p, t-1				6.9	0.34

[a] A minus sign in parentheses, "(-)", indicates that the sign of the simple correlation coeficient between the preceding variable and SS_e is negative; if no sign is shown, the correlation is positive.

EFFECTS OF PROCESS PARAMETERS

Correlations between influent, operational and environmental process parameters and effluent BOD and SS concentrations were studied using multiple linear regression analysis. Regression equations using combinations of various process parameters as independent variables were constructed in a stepwise manner. All data used in constructing the regression equations were normalized so that the regression coefficients would be easier to interpret.

The regression analysis was conducted in two parts. First, regression equations were developed using simple linear combinations of untransformed variables. In the second part, various transformations and combinations of the independent variables used in the first part were included in the regression analysis. In both steps, only those independent parameters that contributed materially to the significance of the regression equation were included in the final form of the equation.

Part I - Simple Combinations of Independent Parameters

The results of the first stage regression analysis are presented in Tables 12a and 12b. All of the regression equations are significant at the 0.1 percent level but the coefficients of determination, R^2 for several of the equations are low and indicate rather weak, though significant, correlations between the combinations of independent parameters and the dependent variables. An asterisk (*) is used to indicate that the sign for the indicated variable is opposite the sign of the simple correlation coefficient between that variable and the dependent variable of the regression.

It may be concluded from the results in Table 12a that primary effluent BOD$_5$ concentration is, in general, highly correlated with effluent BOD$_5$ concentration. Flow rate, influent BOD$_5$ concentration, and the reciprocal of influent wastewater temperature are also highly correlated with effluent BOD$_5$ for most plants where data were available. Results for the other independent parameters are inconclusive either because there was little correlation, the correlation varied from plant to plant or because data for that variable were not available from most plants.

The regression equation for effluent SS generally have somewhat lower coefficients of determination, indicating that effluent SS concentrations are not correlated to other process parameters as highly as are effluent BOD$_5$ concentrations. In Table 12b, it can be seen that primary effluent SS, the inverse, of influent water temperature, flow rate and influent SS are the independent parameters most highly correlated with effluent suspended solids concentrations.

Part II - Combinations of Transformed Independent Variables

The results of the part II regression analysis are presented in Tables 13a and 13b. The independent variables are listed in the order in which each variable was entered into each regression equation. An asterisk (*) is used to indicate that the sign on the indicated independent variable is opposite the sign of the simple correlation coefficient between that variable and dependent variable of the regression.

The R^2 values for these regression equations are generally higher than for the regressions in Part I, but the increase is less than 0.05 for about one-third of the data sets. From an examination of the independent variables, it can be seen that the parameters which were significant in the simpler regression equations are also important in the Part II regressions.

Some of the transformed terms were significant while others were not. Lagged values of influent and primary effluent concentrations, as well as loading terms (flow rate times concentration), were included in over one half of the Part II regression equations. The inclusion of these loading and lagged terms appears to increase significantly the coefficients of determination for several of the regression equations. The squares of BOD_i, BOD_p, SS_i and SS_p were also included in many of the equations. Other transformed terms were not consistently important, although combinations of air and water temperature, either AVTEMP or DTEMP, were included in the equations for effluent BOD_5 for all three plants where air temperature data were available.

VARIATION CAUSED BY MEASUREMENT ERROR

The estimates of variation caused by measurement error are given in Table 14 for effluent concentration data from 11 trickling filter plants. Variation for the SS test is greater than for the BOD_5 test for every plant. The standard deviation of the analytical tests is estimated to be between 1.8 g/m^3 and 7.5 g/m^3 for BOD_5 and between 5.3 g/m^3 and 10.7 g/m^3 for SS. Although these results are only approximate, it is apparent that a significant percentage of the variation in effluent concentration data is due to the limited precision of the BOD_5 and SS tests.

BEGIN FIRST LINE OF TEXT HERE
CENTER OF PAGE
TOP OF IMAGE AREA
DROPPED HEAD: BEGIN SECTIONS HERE

TABLE 14

VARIATION CAUSED BY MEASUREMENT ERROR

	Effluent BOD_5		Effluent Suspended Solids	
Plant Number	Estimated Percent of Variation Caused by Measurement Error	Standard Deviation g/m^3	Estimated Percent of Variation Caused by Measurement Error	Standard Deviation g/m^3
1	10	4.7	11	10.5
2	7	1.9	23	6.4
3	6	1.8	14	6.3
4	13	7.5	51	10.7
5	15	4.4	44	5.9
6	36	4.0	78	5.3
7	5	3.6	21	6.8
8	78	5.8	37	8.2
9	12	6.9	27	9.1
10	28	3.3	15	5.5
11	6	3.0	37	5.4

SECTION 6

DISCUSSION OF RESULTS

EFFLUENT DISTRIBUTIONS

For any given parameter, the distribution which was best for describing effluent data was generally different for each treatment plant. Three distributions, the two-parameters empirical, log-normal, and gamma, were generally comparable and could be used to describe effluent data better than the other two distributions tested. These three distributions can be used to model effluent distributions adequately, but no single distribution, empirical or theoretical, is generally best for modeling effluent distribution from all trickling filter plants.

Literature Comparison

The general characteristics of trickling filter effluent concentration distributions found in this study are in a agreement with the results of previous studies on trickling filter and activated sludge plants. Other workers have determined that effluent distributions vary among different treatment plants (20,44) and that daily effluent concentration data usually are positively skewed. (1,38,43) As was previously shown for activated slude plants, an empirical model consisting of a series of linear regression equations can be developed to model effluent distributions. (21,39)

Relative Merits of Different Distributions

All of the distributions, except the one-parameter model, fit trickling filter effluent data with roughly the same degree of accuracy. Therefore, the choice of distribution to use in any particular application can be made on the basis of the relative ease with which the different distributions are calculated. Even the one parameter model can be used to obtain a reasonable, rough estimate of effluent distributions and may be preferable in some applications.

The advantages of using the one-parameter empirical model is that the distribution standard deviation does not have to be calculated or estimated. When the standard deviation must be estimated, the one-parameter model is as accurate for predicting effluent distributions as any of the two-parameter models, unless an unusually good estimate of the standard deviation is available.

Percentiles can be calculated easily for any of the five distributions once the parameters of the distribution are specified. The arithmetic mean and (when required) the standard deviation are sufficient to specify any of the distributions except the log-normal. To use the log-normal distribution, the mean and standard deviation of the logarithms of the data are required, and these are not the same as the logarithms of the arithmetic mean and standard deviation.

If a desired percentile value is specified and (when required) the standard deviation is estimated, any of the five distributions can be used to compute the implied mean value of the distribution. If the gamma distribution is used, an iterative procedure, which is time-consuming unless performed on a computer, must be used to calculate the distribution mean.

Other disadvantages of using the gamma distribution is that only relatively incomplete tables of gamma distribution values are available, and gamma percentiles are difficult to compute. In contrast, complete tables of the normal distribution are available. With some simple conversions, normal tables can be used for the log-normal distribution as well. Of course, tables are not available for the two empirical distribution models, but the calculations involved in using these distributions are simple.

Effluent Standards

The distributions of effluent concentrations must be known if reasonable and consistent effluent standards are to be set. A reasonable standard does not require a level of treatment which cannot be achieved with available technology. Two standards are consistent if neither standard requires a significantly higher level of treatment than the other.

Hovey et al (21) has shown that EPA secondary standards are both reasonable and consistent when applied to activated sludge plants. A similar analysis can be performed for trickling filters using the one-parameter models developed in this study.

Federal EPA secondary standards require that the maximum (100 %ile) continuous 7-day average effluent BOD_5 and SS concentrations be equal to or less than 45 milligrams per liter and that the maximum 30-day average effluent BOD and SS concentrations be less than or equal to 30 g/m^3. The one-parameter regression equations using the appropriate coefficients found in Appendix A for the appropriate 100%ile values are:

$$100\%\text{ile (7-day } BOD_e) = 1.26 \text{ (mean } BOD_e) + 17.03 \ g/m^3 \quad (30)$$

$$100\%\text{ile (30-day } BOD_e) = 1.31 \text{ (mean } BOD_e) + 7.08 \ g/m^3 \quad (31)$$

$$100\%\text{ile (7-day } SS_e) = 1.50 \text{ (mean } SS_e) + 13.36 \ g/m^3 \quad (32)$$

$$100\%\text{ile (30-day } SS_e) = 1.42 \text{ (mean } SS_e) + 3.63 \ g/m^3 \quad (33)$$

Substituting the appropriate %ile values and solving for the required mean values yields:

for 7-day BOD_e implied mean $BOD_e = 22.3 \ g/m^3$

for 30-day BOD_e implied mean $BOD_e = 17.5 \ g/m^3$

for 7-day SS_e implied mean $SS_e = 21.1 \ g/m^3$

for 30-day SS_e implied mean $SS_e = 18.6 \ g/m^3$

Of the 11 treatment plants in this study, two had a mean effluent BOD$_5$ concentration equal to or less than 17.5 g/m^3 and three had a mean effluent SS concentration equal to or less than 18.6 g/m^3. EPA secondary standards were attained by a relative minority of trickling filter plants evaluated in this study.

The implied mean effluent BOD$_5$ concentration for the 7-day average standards is 4.8 g/m^3 higher than the implied mean for the 30-day standard. For effluent SS concentrations, the implied mean for the 7-day standard is 2.5 g/m^3 higher than for the 30-day standard. From these results, the conclusion may be made that the 30-day effluent BOD$_5$ standard is significantly more restrictive than the 7-day standard and that the 30-day effluent SS standard is somewhat more restrictive than the 7-day SS standard. EPA secondary standards are therefore not consistent when applied to trickling filter treatment plants.

RANDOMNESS OF EFFLUENT DATA

It is clear from the results of the Runs tests and the tests for serial correlation that trickling filter effluent BOD$_5$ and SS data are highly non-random. These results are in agreement with the work of Townshend. (45) Also, it appears that for each plant studied, the degree of randomness for effluent SS data is either equal to or greater than the degree of randomness for effluent BOD$_5$ data.

Distribution functions normally are used to make statements concerning the likelihood of the occurence of some event, irrespective of the conditions under which the event occurs. The time at which the event occurs or the history of any previous events is not specified because it is assumed that the likelihood of the event is independent of these factors. For example, the probability of throwing a seven when rolling two dice is independent of when the dice are rolled or of the results of any previous rolls.

Effluent concentration data in this study are not randomly generated and are thus not independent of time and preceding effluent concentrations. The probability that an effluent BOD$_5$ (or SS) concentration will be exceeded on any one day is dependent on which particular day is of interest. Information concerning the conditions under which events occur is not included in distribution functions, and hence, not statement concerning the probable level of effluent concentrations on any one day can be made based simply on the distribution models considered in this study.

Distribution models for effluent concentrations do provide information on the long-term behavior of trickling filter performance and can be used to estimate the probable frequency of occurrence of some range of effluent values over a one year period.

ANNUAL AND WEEKLY PATTERNS

On the basis of the analysis of variance results and the plots of pooled, normalized data, the following conclusions can be made concerning effluent BOD$_5$ and SS concentrations at the 11 treatment plants studied.

(1) There are significant month to month fluctuations in average effluent concentrations for every treatment plant studied.

(2) One cause of these monthly fluctuations is an annual cycle in effluent concentrations which is evident when data from all 11 treatment plants is pooled. Effluent concentrations tend to be above average in the winter and below average in the summer.

(3) A significant weekly cycle exists in effluent concentration data from some of the treatment plants studied.

(4) When data from nine treatment plants were pooled, weekly cycles were evident for BOD_5 and SS data, but the amplitudes of these cycles were less than half the amplitudes of annual cycles. Effluent BOD and SS concentrations tend to be above average on Wednesdays.

Literature Comparison

Conclusions (1), (2), and (3) above are in agreement with previous work on annual and weekly cycles in effluent data from trickling filter plants. (20,45) Based on this study and work by Popel (38), it appears that monthly average effluent data from trickling filter and activated sludge plants fluctuate significantly. However, activated sludge plant effluent data does not appear to contain seasonally dependent annual cycles such as found in the trickling filter plant data. (45)

Townsend (45) and Hann et al (20) both found seasonally dependent annual cycles in trickling filter effluent data similar to the cycles found in this study. The 11 treatment plants in this study and the 16 trickling filter in the study by Townsend were generally located in the upper midwestern United States or Ontario, Canada, areas where the winters are relatively severe. Hann concluded that there were annual cycles in two of four trickling filters studied; all of which were located in Texas. Therefore, the evidence for the existence of annual cycles in effluent data for trickling filter plants located in colder climates is relatively conclusive, but it remains unclear whether annual cycles are common for plants located in more temperate climates.

The fact that weekly cycles were found in effluent data from some, but not all, treatment plants studied agrees with the conclusions of studies by Hann et al (20) and Popel. (38) Hann concluded that effluent concentrations were generally lower on weekends, while in the present study, effluent concentrations were found to be significantly below average on Sunday only.

Causes of Annual and Weekly Cycles

The annual cycles observed in the pooled effluent data may be caused by seasonal changes in water and air temperature, flow rate, and other environmental factors. It was found in the regression analysis that for most data sets, effluent concentrations were negatively correlated with water and air temperature. Lower winter temperatures reduce biological reaction rates and thus tend to impar process performance. The opposite effect occurs in the summers.

It was also found that effluent concentrations tended to be positively correlated with flow rate. However, this correlation does not help explain the annual cycles of effluent concentrations because effluent concentrations tend to be above average in December through March, but flow rates in the treatment plants studied were above average in March through May.

The weekly cycles found in some effluent BOD and SS data sets are probably caused by weekly cycles in influent BOD and SS concentrations and flow rate. Influent concentrations and flow rate are all positively correlated with effluent concentrations for data from most of the treatment plants studied.

EFFECTS OF PROCESS VARIABLES ON EFFLUENT CONCENTRATIONS

Based on the results of the multiple regression analyses, it can be concluded that influent, operational, and environmental process variables are correlated with effluent BOD_5 and SS concentration data at trickling filter plants. It has not been proven that fluctuations in effluent data are caused by changes in the independent parameters because regression analysis alone cannot be used to prove cause and effect relationships. Correlations between two variables conceivably could be caused by a third, unmeasured variable. however, the relationships found are consistent with current theoretical models of clarifiers and trickling filters and provide strong evidence to support the hypothesis that a significant percentage of effluent variation in trickling filter plants is caused by fluctuations in process loading and temperature.

Comments on Part I of the Regression Analysis

Primary effluent BOD_5 and SS concentrations were more highly correlated to effluent concentrations than were influent concentrations. One reason for this result is that a certain amount of randomness or "noise" is introduced into the influent data series by the sedimentation process. Other treatment processes increase the randomness of the data series too. This cumulative effect of randomizing the incoming data series is probably one reason that the coefficients of determination for the regression equations are not higher. Wastewater treatment processes are stochastic in nature and attempts to explain effluent variability in deterministic terms will explain only a portion of the phenomenon.

Comments on Part II of the Regression Analaysis

Lagged values of influent and primary effluent BOD and SS concentration and loading terms (flow times concentration) were significantly correlated with effluent concentrations for roughly half of the treatment plants studied. Thus, terms of these types can be useful in deterministic models to predict effluent concentrations from process variables.

The correlations of lagged terms with effluent concentrations are especially interesting because the implication is that plant performance can be affected by influent and primary effluent concentrations from the previous day. No mechanistic explanation can be given for this observation. It is possible that the correlations for lagged terms were due to the influence of some third factor.

VARIATION DUE TO MEASUREMENT ERROR

Measurement errors can cause a significant percentage of the total variation in effluent concentrations, especially for SS data. The percentage of the variation in effluent data caused by analytical error was estimated to be greater than 10 percent for BOD_5 and SS concentrations from seven and eleven plants, respectively, and was greater than 20 percent for BOD_5 and SS data from three and eight plants, respectively.

General Comments

The greater variability and randomness found in suspended solids effluent data may in part be due to the fact that measurement errors appear to be somewhat greater for SS than for BOD_5. However, one reason that measurement errors were estimated to be greater for SS data was because of the assumption that duplicate BOD_5 samples were analyzed, but only one SS sample was used. If this assumption is not correct, the estimated measurement errors for SS analysis would be smaller, though they would still be somewhat larger than the estimated errors for BOD_5 data.

Effluent variation due to measurement errors is independent of variation due to processes parameters. Therefore it should be possible to add the percentages of variance explained by the regression analysis and by the estimates of measurement error. The sums of the part II regression coefficients of determination and the estimated percentages of variance due to measurement error are shown in Table 15.

The sum of the percentages of variance explained by the regression models and by measurement error is greater than 40% for all but 3 sets of data. From these results, it can be concluded that a large fraction of effluent variability is caused by variation in process parameters and by measurement error. Becuase three of the sums are greater than 100 percent, the estimates of measurement error are in some cases too high.

Measurement errors do not affect the values of running average data as much as they affect daily data, because the errors tend to cancel out. For example, if the standard deviation of a test is 5.0 g/m^3, the standard deviations of 7-day and 30-day averages for the same parameters would be 1.9 g/m^3 and 0.9 g/m^3, respectively.

ENGINEERING APPLICATIONS

the existence of a seasonally dependent yearly pattern in trickling filter plant performance has possible implications for wastewater treatment planning, design, and regulation. In areas where winter sewage flow rates are especially high and winter temperatures quite low, a treatment process that is less prone to seasonal fluctuations than trickling filters would provide more stable performance. Because trickling filter performance tends to be poorest during winter months, design parameters such as flow rate and temperature should be chosen for winter conditions. Finally, in situations where stream water quality must be maintained, the fact that trickling filter plant performance is poorest in the winter would be ameliorated by higher stream flow rates and higher oxygen saturation concentrations in the winter months.

If effluent concentrations are not sampled every day, the measured weekly and monthly average BOD and SS concentrations could be either higher or lower than

TABLE 15

SUM OF PART II REGRESSION R^2 AND ESTIMATED PERCENTAGE OF VARIANCE DUE TO MEASUREMENT ERRORS

Plant Number	Sum of R^2 and EPV, percent	
	BOD_5	Suspended Solids
1	56	37
2	44	47
3	33	41
4	104	125
5	61	74
6	45	115
7	53	44
8	103	86
9	79	76
10	52	32
11	41	71

would be the case with daily sampling. For example, effluent concentrations tend to be lowest on Sundays, yet this is the day most likely not to be sampled.

Based on the results of this study, it may be concluded that the magnitude of variation in effluent data caused by measurement errors is quite large at many treatment plants. If a treatment plant is operating near its discharge limits, apparent violations of the daily and 7-day average concentration limits can be caused by analytical error alone.

To reduce the possibility of discharge violations caused by measurement error, more replicates of each sample can be analyzed. Also, sampling and laboratory techniques should be checked to see that all sources of sampling and analytical errors are minimized.

SECTION 7

RELIABILITY AND STABILITY OF TRICKLING FILTER PROCESSES

The uncertainties associated with plant design and operation should be recognized in the design of trickling filters. The best way to accomplish this is incorporate stochastic concepts and procedures in the design process. To produce an effluent of high quality and to meet the effluent discharge requirements at minimum cost, design engineers must be able to estimate the average effluent quality and its variations for a given treatment process. Uncertainties and their significance on process performance can be analyzed systematically using methods of probability. A probabilistic approach for design can be used to provide a consistent basis for the analysis of uncertainty and a theoretical basis for the analysis of performance and reliability.

In a previous study by Niku et al (32,33), the reliability and stability of activated sludge processes have been examined. In this study, an attempt will be made to determine if the reliability and stability models developed for activated sludge processes can be applied to the trickling filter processes, also. For more details about the procedures of the model development of reliability and stability of activated sludge processes, the reader is referred to the references 32 and 33.

EFFLUENT QUALITY DISTRIBUTION OF TRICKLING FILTERS

In the previous sections of this paper, the effluent BOD$_5$ and suspended solids concentration distributions of trickling filter processes were examined. It was found that except in one plant, the skewness coefficient is always positive (Table 6A) indicating that generally the data were not symmetrical and were skewed to the right of the most frequent values. Several theoretical and empirical distributions were tested in order to characterize the effluent data of trickling filters. It was found that the log-normal distribution fits the effluent BOD$_5$ and SS data, and this distribution can not be rejected for at least half of the plants data at significance levels of one percent (Tables 8A and 8B). It should be noted that other distributions such as gamma may fit the effluent data as well, but for simplicity and because of availability of statistical tables for log-normal distribution, the log-normal distribution is preferred. The similar effluent characteristics of trickling filters and activated sludge processes lead tot he conclusion that the reliability model already developed for activated sludge processes may be used for reliability estimation of trickling filters.

RELIABILITY CONCEPT - BACKGROUND

The reliability measures are expressed in probability terms and are defined as the probability of adequate performance, i.e., the percent of the time that effluent concentrations meet requirements. A stochastic model and design tables and graphs have been developed in a previous study (32) predicting achievable effluent BOD_5 and SS concentrations based on a proposed coefficient of reliability (COR). The COR relates mean constituent values to the standard and is achieved on a probability basis and the log-normal distribution assumption. The COR was defined mathematically as: (32)

$$COR = (V_x^2 + 1)^{1/2} \exp\{-Z_{1-\alpha}[Ln(V_x^2 + 1)]^{1/2}\}$$

where: $V_x = \frac{S_x}{m_x}$ is coefficient of variation, equal to standard deviation (S_x) divided by mean (m_x) of the data.

$Z_{1-\alpha}$ = F(Z) where Z is normally distributed with mean zero and variance one and F(Z) may be obtained from the standard normal variate tables. (5)

α is the probability of failure of meeting standards.

The mean constituent value, m_x (i.e. the design value), is obtained from

$$m_x = (COR)X_s$$

where X_s = a fixed standard (e.g. $X_s = 30$ g/m^3).

Values of the COR for effluent concentrations for different coefficients of variation (V_x) at different levels of reliability are given in Table 16. Values of the COR given in Table 16, may be used in design of a trickling filter process expected to perform at a certain reliability. For example, to have 95% reliability that the effluent BOD_5 concentration of a trickling filter process will be equal to or less than a certain standard X_s, when the coefficient of variation is estimated to be $V_x = 0.50$, the plant should be designed for the mean value equal to or less than $0.51 X_s$. (If $X_s = 30$ g/m^3; then average BOD = 0.51 x 30 = 15.3 g/m^3 or less.)

The coefficient of variation for effluent concentration should be estimated from past experience and on reasonable expectation of performance. Values of the coefficient of variation from other similar treatment plants such as those given in Table 6a may be used as a guidance. Based on cursory examination, coefficient of variations of 0.50 for BOD_5 and 0.55 for SS concentration would appear to be suitable values, however.

The fractional reliability for the plants under operation is presented in Table 17, as a function of V_x and the mean to standard concentration ratio (COR). Results from Table 17 may be used to estimate the reliability of a trickling filter process under operation if the mean and the standard deviation for the plant is known. For example, the reliability of a plant operating with a mean value of $0.8X_s$ and a V_x = 0.50 is 76 percent. In other words, in a long-run operation, about 24 percent of the observed effluent variables would exceed 30 g/m^3 in a plant operating with mean value of 24 g/m^3 and standard deviation of 12 g/m^3 $(V_x = 0.50)$.

TABLE 16

COR* as a Function of V_x and Reliability

Based on True Log-Normal Distribution

Reliability	50%	80%	90%	92%	95%	98%	99%	99.9%
V_x								
0.3	1.04	0.81	0.71	0.69	0.64	0.57	0.53	0.42
0.4	1.08	0.78	0.66	0.63	0.57	0.49	0.44	0.33
0.5	1.12	0.75	0.61	0.58	0.51	0.42	0.37	0.26
0.6	1.17	0.73	0.57	0.54	0.47	0.37	0.32	0.21
0.7	1.22	0.72	0.54	0.50	0.43	0.33	0.28	0.17
0.8	1.28	0.71	0.52	0.48	0.40	0.30	0.25	0.15
0.9	1.35	0.70	0.50	0.46	0.38	0.28	0.22	0.12
1.0	1.41	0.70	0.49	0.44	0.36	0.26	0.20	0.11
1.2	1.56	0.70	0.46	0.41	0.33	0.22	0.17	0.08
1.5	1.80	0.72	0.45	0.39	0.30	0.19	0.14	0.06

$$* \quad COR = \frac{m_x}{X_s} = (V_x^2 + 1)^{1/2} \exp \quad -Z_{1-\alpha} \quad \mathrm{Ln}(V_x^2 + 1)^{1/2}$$

TABLE 17

Reliability as a function of Normalized Mean, and V_x of Effluent Concentration

$\frac{m_x}{X_s}$	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0	1.5	2.0
V_x												
0.3	.9999	.9999	.9999	0.9995	0.9939	0.9706	0.9134	0.8179	0.6932	0.5583	0.1083	0.0136
0.4	.9999	.9999	0.9995	0.9950	0.9767	0.9353	0.8681	0.7799	0.6793	0.5763	0.1950	0.0542
0.5	.9999	.9999	0.9973	0.9852	0.9557	0.9062	0.8391	0.7607	0.6770	0.5934	0.2669	0.1093
0.6	.9999	0.9993	0.9928	0.9732	0.9366	0.8846	0.8212	0.7516	0.6798	0.6092	0.3251	0.1660
0.7	.9999	0.9979	0.9869	0.9610	0.9212	0.8675	0.8107	0.7483	0.6853	0.6239	0.3721	0.2171
0.8	.9999	0.9959	0.9805	0.9509	0.9087	0.8594	0.8047	0.7482	0.6920	0.6373	0.4110	0.2632
0.9	.9996	0.9932	0.9743	0.9423	0.9006	0.8527	0.8018	0.7539	0.6992	0.6499	0.4439	0.3033
1.0	.9993	0.9905	0.9688	0.9353	0.8941	0.8485	0.8009	0.7530	0.7065	0.6614	0.4718	0.3398
1.2	.9982	0.9851	0.9597	0.9254	0.8860	0.8445	0.8023	0.7604	0.7201	0.6816	0.4828	0.3967
1.5	.9961	0.9785	0.9507	0.9172	0.8812	0.8446	0.7754	0.7728	0.7389	0.7075	0.5675	0.4620

BEGIN FIRST LINE OF TEXT HERE
DROPPED HEAD BEGIN SECTIONS HERE
BEGIN LAST LINE OF TEXT
EPA Form 2350-4 (4-80)
TYPING GUIDE SHEET

The results of Table 16 and 17 are presented in graphically in Figure 7. This figure can be used in both estimation of the mean value in design of a process or prediction of mean effluent quality in a plant under operation.

VERIFICATION OF THE RESULTS

To support the validity of the reliability model in prediction of effluent quality performance from trickling filters, the percent of the time that effluent concentration exceeded 30 g/m^3 was computed for one year of data in all 11 plants (Table 18). This percent of exceedance was also predicted using the reliability model based on the log-normal assumption. The prediction values are very comparable with the measured values.

To determine the validity of the theoretical results of the COR, Table 17 was reconstructed based on the one year of operational data of all 11 plants (results are not shown). The same procedures as used for similar validation studies on activated sludge processes (32) were used. Percentile values of exceedance for effluent BOD_5 and SS from one year of daily operational data have been computed to present an empirical distribution of effluent concentration without considering whether it follows any classical distribution function. The results of COR based on pooled plant data were highly comparable to the results of COR based on the log-normal distribution assumption.

STABILITY OF TRICKLING FILTER PLANTS

Niku et al (33) has defined stability as a measure of adherence to the annual mean constituent concentration. In the same study it was found that the "standard deviation" is the most appropriate measure of stability in wastewater treatment plants. Through the statistical analysis of the stability of 43 activated sludge wastewater treatment processes it was found that plants with an annual standard deviation of less than 10 g/m^3 are statistically stable, whereas those with standard deviation of greater than 10 g/m^3 are statistically unstable.

In Figures 8 and 9, the mean, standard deviation, and range of effluent BOD_5 and SS data for all 11 trickling filter plants are shown. Examination of these descriptive statistics results in the conclusion that the same stability cut-off point can be used as 10 g/m^3 in activated sludge processes. That is, a distinct difference exists between the statistical characteristics of the plants operating below and above the standard deviation of 10 g/m^3. The maximum daily concentrations of effluent BOD_5 and SS in stable plants with standard deviation or less than or equal 10 g/m^3 are usually less than 70 g/m^3 whereas in unstable plants maximum concentrations are usually greater than 90 g/m^3 and 80 g/m^3, respectivley.

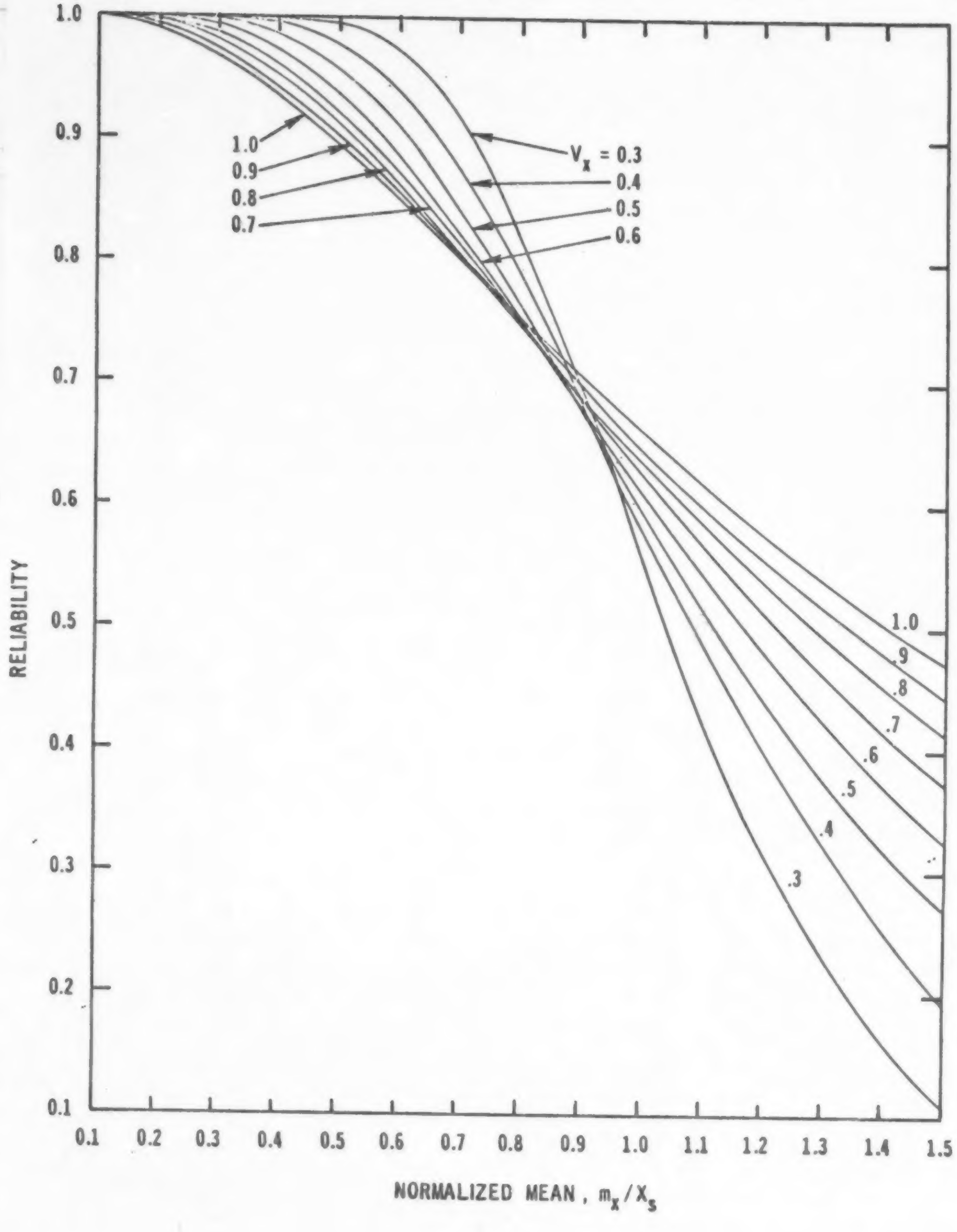


Figure 7. Reliability versus normalized mean for different coefficients of variation.

TABLE 18

COMPARISON OF PREDICTED AND MEASURED PERFORMANCE OF TRICKLING FILTER PROCESSES

	BOD				SS			
			%time $\leq$ 30 g/m^3				%time $\leq$ g/m^3	
Plant Number	V_X	X/30	Predicted	Measured	V_X	X/30	Predicted	Measured
1	0.46	1.11	49.3	49.6	0.59	1.75	22.6	19.4
2	0.67	0.36	97.6	97.5	0.62	0.72	80.6	78.1
3	0.75	0.34	97.4	98.3	0.79	0.70	80.5	76.2
4	0.36	1.94	4.3	8.0	0.27	1.83	1.6	5.6
5	0.38	0.94	63.8	57.9	0.48	0.61	90.5	93.1
6	0.25	0.90	70.9	72.6	0.40	0.50	97.7	98.1
7	0.73	0.77	76.6	75.5	0.63	0.80	75.0	70.7
8	0.15	1.44	2.1	1.1	0.40	1.13	45.0	45.8
9	0.39	1.70	11.1	17.8	0.43	1.37	28.8	28.5
10	0.30	0.70	91.4	97.6	0.61	0.79	75.8	72.8
11	0.65	0.61	87.1	88.2	0.55	0.54	92.7	93.4

Predicted percentages represent areas under the normal probability curve F(z) where z is equal to

$$z = \frac{\ln\left[\frac{mx}{X_s}\,(V_x^2 + 1)^{-1/2}\right]}{\left[L_n\,(V_x^2 + 1)\right]^{1/2}}$$

. See reference 32 for details.

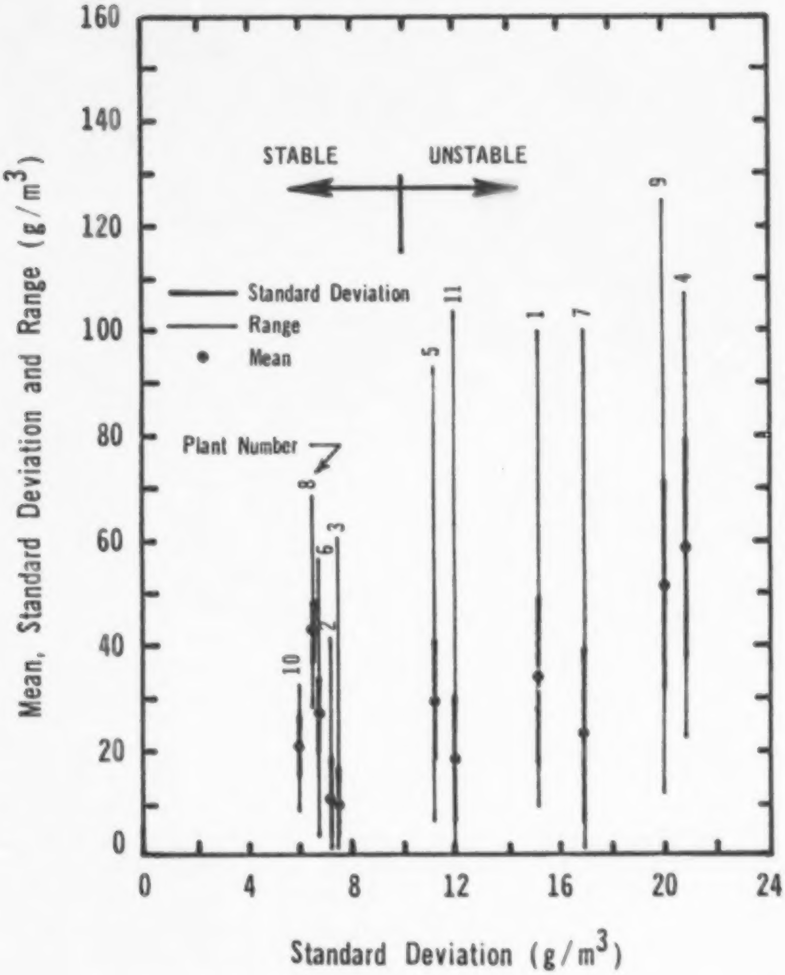


Figure 8. Variability of effluent BOD as a function of standard deviation.

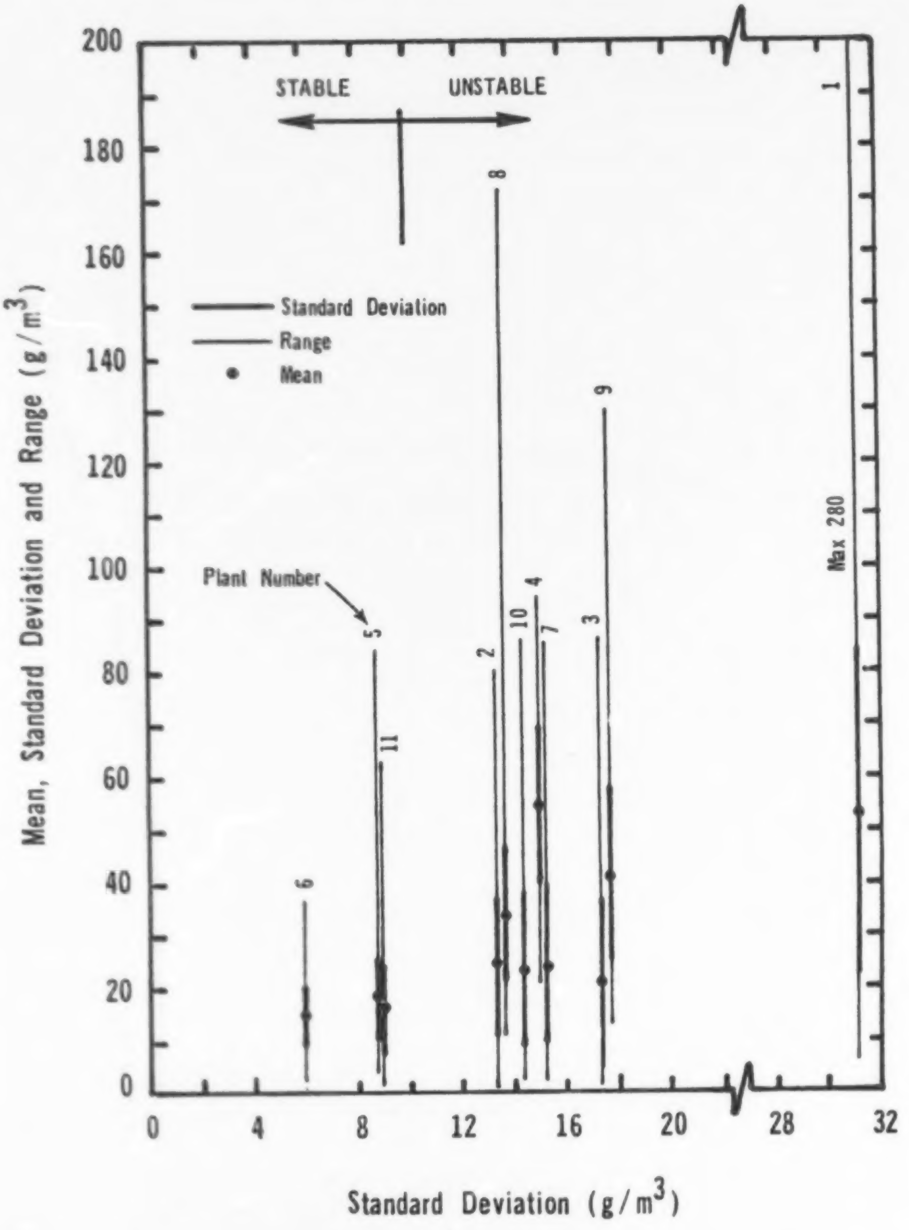


Figure 9. Variability of effluent suspended solids concentration as a function of standard deviation.

SECTION 8

RECOMMENDATIONS

Because of the scope of this study, it was not possible to examine some of the aspects of trickling filter plant effluent variability in great detail. Some questions concerning annual cycles in the data and the effects of certain parameters on effluent concentrations could not be answered because most of the data sets were only 1 year in length and did not contain data for all the parameters of interest. Also, as in any research, finding answers to one question invariably raises more. To conclude this discussion of results, suggestions for further research are presented below.

EFFLUENT DISTRIBUTIONS

Even though the shape of the effluent distributions change somewhat from plant to plant, most effluent distributions can be described adequately using any of the two-parameter distribution models used in this study. The biggest disadvantage of using those models to predict effluent distributions is that the standard deviation of the distribution based on general characteristics of treatment plants such as mean flow, process configuration and ratio of average flow to design flow, would make it possible to predict effluent distributions with greater accuracy.

Conclusions derived from this part of the study should be substantiated for a larger data base. Of particular interest would be to determine if the distributions of effluent data at a treatment plant change from year to year.

ANNUAL AND WEEKLY CYCLES IN EFFLUENT DATA

No conclusions were made concerning annual cycles in the data from individual treatment plants because only one year of data was available from each plant studied. Also, most of the plants in this study were located in the northern part of the U.S., and hence, the observations concerning annual cycles in pooled data were restricted to treatment plants located in areas with cold winters. To compensate for these shortcomings in the present study, several years of effluent data from treatment plants located throughout the country should be tested for annual cycles.

While annual and weekly cycles are two of the major non-random components of trickling filter effluent data, part of the non-random remains unexplained. For example, part of the fluctuations in average monthly performance are due to seasonally dependent annual cycle. However, much of these fluctuations are either part of some other, unkown cycle or representative of non-cyclical changes in performance from month to month.

EPA Form 2350-4 (4-80)
(PREVIOUSLY CIN. EPA FORM 247)
TYPING GUIDE SHEET

The methods used in this study to examine cycles in effluent data were relatively unsophisticated. A more complete characterization of the non-random component of trickling filter effluent data should be made using sophisticated time-series methods such as spectral analysis or Box-Jenkins methods. Better understanding of the time-dependent behavior of effluent data would improve prediction of effluent distributions and lend insight into treatment process dynamics.

EFFECTS OF PROCESS VARIABLES ON EFFLUENT CONCENTRATIONS

A significant percentage of the variation observed in trickling filter effluent data has been explained by using linear regression analysis. Much of the variation remains unexplained, however. Part of the variation in effluent data is stochastic in nature and cannot be modelled in deterministic terms. Investigations into deterministic causes of trickling filter effluent variability (beyond the scope of the present study) are needed. For example, few plants provided data for three of the process parameters studied in the regression analysis: recycle ratio, air temperature and influent wastewater pH. These parameters, especially recycle ratio and air temperature, should be studied more thoroughly to see if they affect effluent variability. Other possible causes of effluent variability that could be studied are discussed in the section on measurement errors.

From the results of the regression analyses, no conclusion can be made as to whether the independent parameters are correlated with the day to day fluctuations in effluent concentrations, or are correlated with both. If the correlations are primarily between short-term fluctuations, then changes in the independent parameters cannot be used to explain the annual cycles in effluent data. If on the other hand, the correlations are only between long-term changes in the independent and dependent variables, then it would be likely that short-term fluctuations are primarily due to random processes.

The question posed above could be studied by subtracting 30-day running averages from each daily effluent concentration and performing two new regressions like the ones in this study but with the remainders of the subtraction and the 3-day averages as the new dependent variables. The residuals from subtracting the 30-day averages from the daily values represent the daily fluctuations and the 30-day averages represent the long-term trends. Because the residuals and running averages are uncorrelated, the two regressions can be used to determine how much of the correlation between process parameters and effluent data is related to daily fluctuations and how much is related to long-term trends. This approach should be considered in future studies investigating trickling filter efficient variability.

MEASUREMENT ERRORS

Measurement error estimates presented in this study were based on a number of assumptions and simplifications. Better estimates of the magnitude of variation caused by measurement error could be made if some of these simplifications were not made.

The precision of the BOD_5 and SS tests should be determined for five or six different concentrations between 5 g/m^3 and 100 g/m^3 so that the precision of these tests is known for typical values of effluent data. Currently, documentation in this area is limited.

The studies on BOD_5 and SS measurement precision found in the literature were conducted by sending identical samples to several different laboratories to be analyzed. At any one lab, the precision of these tests may be better or worse than the precision of several labs. Variation in the precision of analytical procedures among different laboratories should be examined.

Two causes of effluent variability in trickling filters have been investigated in this study: variation in process parameters and measurement error. There are a number of possible causes of the variation which was not explained in these two analyses. The focus of further study in this area should be on the following factors:

- inaccuracies in the hypothesized functional relationships between parameters.
- biological cycles independent of any measured parameter
- sampling errors
- variations in the composition of influent constituents not included in the regression analyses
- changes or malfunctions in the operation of treatment processes not studied in the regression analysis.

REFERENCES

1. Adams, B.J., and Gemmell, R.S., An Analysis of Individual and Grouped Performance of Regionally-Related Wastewater Treatment Plants, Urban Systems Engineering Center, Northwestern University (1973).

2. Adams, B.J., and Gemmell R.S., "Mean Estimate Deficiencies in Water Quality Studies." Jour. Hydraulics Div., Proc. Amer. Soc. Civil Eng., 101, HY7, 989 (1975).

3. Ballinger, D.G., and Lishka, R.J., "Reliability and Precision of BOD and COD Determinations." Jour. Water Poll. Control Fed., 34, 5, 470 (1962).

4. Beck, M.B., "Dynamic Modelling and Control Applications in Water Quality Maintenance." Water Research, 10, 7, 575 (1976).

5. Benjamin, J.R., and Cornell, A.C., Probability, Statistics and Decision for Civil Engineers, McGraw-Hill Book Company (1970).

6. Berthouex, P.M., and Polkowski, L.B., "Optimum Waste Treatment Plant Design Under Uncertainty." Jour. Water Poll. Control Fed., 42, 9, 1589 (1970).

7. Berthouex, P.M., "Some Historical Statistics Related to Future Standards." Jour Envr. Eng. Div., Proc. Amer. Soc. Civil Eng., 100, EE2, 423 (1974).

8. Berthouex, P.M., et al., "Modeling Sewage Treatment Plant Input BOD Data." Jour. Envr. Eng. Div., Proc. Amer. Soc. Civil Engr., 101, EE1, 127 (1975).

9. Berthouex, P.M., et al., "The Use of Stochastic Models in the Interpretation of Historical Data from Sewage Treatment Plants." Water Research. 10, 8, 689 (1976).

10. Box, G.E.P., and Jenkins, G.M., Time Series Analysis Forecasting and Control, Holden-Day (1970).

11. Climatological Data for the United States by Section, Iowa, 87, 1-12. National Climatic Center (1970).

12. Climatological Data for the United States by Section, Michigan, 91, 112. National Climatic Center (1976).

13. Climatological Data for the United States by Section, Michigan, 92, 112. National Climatic Center ((1977).

14. David, M.H., et al., "Statistical Study of Phosphorus Removal in Wisconsin." Jour. Envr. Eng. Div., Proc. Amer. Soc. Civil Engr., 102, EE2, 433 (1976).

15. DiToro, D.M., "Statistical Design of Equalization Basins." Jour. Envr. Eng. Div., Proc. Amer. Soc. Civil Engr., 100, EE6, 917 (1975).

16. Eckenfelder, W.W., Jr., and Adams, C.E., Jr., "Design and Economics of Joint Wastewater Treatment." Jour. San. Eng. Div., Proc. Amer. Soc. Civil Engr., 98, SA1, 153 (1972).

17. Eckenfelder, W.W., Jr. and Roth, J.A., "Effluent Variability as Related to Developing Performance Standards." Progress in Water Technology, 8, 1, 1 (1976).

18. Englande, A.J., Jr., and Mojgani, P., "Estimating Effluent Variability from Complete-Mix Activated Sludge Treatment Systems." Progress in Water Technology, 8, 1, 27 (1976).

19. Fair, G.M., and Geyer, J.C., Elements of Water Supply and Waste-Water Disposal, John Wiley and Sons, Inc. (1965).

20. Hann R.W., et al., "Evaluation of Factors Effecting Discharge Quality Variation, Final Report For the Texas Water Quality Board. Environmental Engineering Division, Civil Engineering Department, Texas A & M University, Austin, Texas (1972).

21. Hovey, W.H., Schroeder, E.D., and Tchobanoglous, G., Optimal Size of Regional Wastewater Treatment Plants, California Water Resources Center, Contribution # 161. University of California, Davis, California (1977).

22. Howell, J.A., and Atkinson, B., "Sloughing of Microbial Films in Trickling Filters." Water Research, 10, 4, 307 (1976).

23. Komolrit, K., and Gaudy, A.F., Jr., "Biochemical Response of Continuous Flow Activated Sludge Processes to Qualitative Shock Loading." Jour. Water Poll. Control Fed., 38, 1, 85 (1966).

24. Lin, K.C., and Heinke, G.W., "Variability of Temperature and Other Process Parameters, A Time Series Analysis of Activated Sludge Data." Progress in Water Technology, 9, 2, 347 (1977).

25. Loucks, D.P., "Risk Evaluation in Sewage Treatment Plant Design." Jour. San. Eng. Div., Proc. Amer. Soc. Civil Engr., 93, SA1, 25 (1967).

26. McLellan, J.C., and Busch, A.W., "Hydraulic and Process Aspects of Reactor Design - Response to Variations in Loading." Proceedings, 24th Industrial Waste Conference, Purdue University, Lafayette, Indiana (1969).

27. Mendenhall, W., and Schaeffer, R.L., Mathematical Statistics with Applications, Duxbury Press (1973).

28. Metcalf and Eddy, Inc., Wastewater Engineering: Collection, Treatment, Disposal, McGraw-Hill Book Company (1972).

29. Methods for Chemical Analysis of Water and Wastes, U.S. Environmental Protection Agency (1974).

30. Mistry, K.J., and Himmelblau, D.M., "Stochastic Analysis of Trickling Filters." Jour. Envr. Eng. Div., Proc. Amer. Soc. Civil Engr., 101, EE3, 333 (1975).

31. Nie, N.H., et al., Statistical Package for the Social Sciences, Second Edition, McGraw-Hill Book Company (1975).

32. Niku, S., Schroeder, E.D., and Samaniego, F.J., "Performance of Activated Sludge Processes and Reliability Based Design." Jour. Water Poll. Control Fed., 51, 2841 (1979).

33. Niku, S., et al., Performance of Biological Wastewater Treatment Plants, Reliability, Stability, and Variability, Report to U.S. Environmental Protection Agency Grant No. R805097-01, (1981).

34. Norton, D.W., An Empirical Investigation of the Effects of Nonnormality and Heterogeneity Upon the F-test of Analysis of Variance, Ph.D. thesis, University of Iowa (1952).

35. Novotny, V., Englande, A.J., and Mojgani, P., "Effluent Variability Estimation for Complete Mix Activated Sludge Treatment Systems." Water Resources, 10, 8, 699 (1976).

36. Novotny, V., and Stein, R.M., "Equalization of Time Variable Waste Loads." Jour. Envr. Eng. Div., Proc. Amer. Soc. Civil Engr., 102, EE3, 613 (1976).

37. Ott, C.R., and Bogan, R.H., "Theoretical Analysis of Activated Sludge Dynamics." Jour. San. Eng. Div., Proc. Amer. Soc. Civil Engr., 97, SA1, 1 (1971).

38. Popel, H.J., "A Concept for Realistic Effluent Standards." Progress in Water Technology, 8, 1, 69 (1976).

39. Porter, K.S., and Boon, A.G., "Cost of Treatment of Wastewater with Particular Reference to the River System of the Trent Area." The Institute of Water Pollution Control, Symposium on the Trent Research Programme (1971).

40. Porter, K.S., "Percentile Mean Relation for Effluent Assessment." Jour. Envr. Eng. Div., Proc. Amer. Soc. Civil Engr., 101, EE3, 473 (1975).

41. Schroeder, E.D., Water and Wastewater Treatment, McGraw-Hill Book Company, New York, (1977).

42. Standard Methods for the Examination of Water and Wastewater, 14th Edition. APHA, AWWA, and WPCF (1976).

43. Study of Discharge Requirements and Effluent Variability, Sacramento Area Consultants. A Report for the Sacramento Regional County Sanitation District (1975).

EPA Form 2360-4 (4-80)
(PREVIOUSLY CIN. EPA FORM 367)

TYPING GUIDE SHEET

44. Thomann, R.V., "Variability of Wastewater Treatment Plant Performance." Jour. San. Eng. Div., Proc. Amer. Soc. Civil Engr., 96, SA3, 819 (1970).

45. Townshend, A.R., "Statistical Analysis of the Effluent Quality of Biological Sewage Treatment Processes." Proceedings, the Third Canadian Symposium on Water Pollution Research, Toronto, Canada, 272 (1968).

46. VonJeszenszky, T., and Dunn, I.J., "Dynamic Modelling and Control Simulation of a Biological Wastewater Treatment Process." Water Research, 10, 5, 461 (1976).

47. Wheatland, A.B., "Statistical Expression of Effluent Quality Standards." Water Research, 6, 339 (1972).

BEGIN FIRST LINE OF TEXT HERE
CENTER OF PAGE
TOP OF IMAGE AREA
DROPPED HEAD, BEGIN SECTIONS HERE
6 1/2"
9 1/8"
BEGIN LAST LINE OF TEXT
3/8"
PAGE NUMBER
BOTTOM OF IMAGE AREA OUTSIDE DIMENSION FOR TABLES AND ILLUSTRATIONS
EPA Form 2350-4 (4-80)
(PREVIOUSLY CIN. EPA FORM 257)
TYPING GUIDE SHEET

APPENDIX A

COEFFICIENTS AND STATISTICAL RESULTS FOR TWO-PARAMETER MODELS

TABLE A-1

COEFFICIENTS AND STATISTICAL RESULTS FOR ONE- AND TWO-PARAMETER EMPIRICAL DISTRIBUTION MODELS DAILY EFFLUENT SUSPENDED SOLIDS

	One-Parameter Model				Two-Parameter Model				
Percentile	A_1	C_1	r^2	Standard Error	A_2	B_2	C_2	r^2	Standard Error
0	0.38	-4.9	0.63	4.1	0.63	-0.80	-1.0	0.93	1.9
10	0.64	-4.8	0.81	4.4	0.91	-0.85	-0.7	0.97	1.9
20	0.80	-5.4	0.89	4.0	1.05	-0.77	-1.7	0.98	1.8
30	0.86	-3.8	0.95	2.9	1.05	-0.58	-1.1	0.99	1.2
40	0.92	-3.1	0.96	2.6	1.10	-0.54	-0.5	0.99	0.7
50	0.97	-1.2	0.98	2.0	1.09	-0.37	0.5	0.99	1.2
60	0.99	1.0	0.99	1.2	1.04	-0.14	1.7	0.99	1.0
70	1.06	2.7	0.99	1.5	1.02	0.10	2.2	0.99	1.5
80	1.17	4.6	0.97	3.1	1.00	0.53	2.0	0.99	2.0
90	1.34	7.1	0.90	6.3	0.93	1.27	1.0	0.99	2.3
95	1.53	8.8	0.87	8.7	0.98	1.72	0.6	0.98	3.4
98	1.72	14.5	0.76	13.7	0.83	2.75	1.3	0.97	5.0
99	2.09	12.9	0.68	20.5	0.72	4.21	-7.3	0.97	6.3
100	3.16	12.6	0.46	49.3	0.52	8.15	-26.5	0.78	33.0

TABLE A-2

COEFFICIENTS AND STATISTICAL RESULTS FOR ONE- AND TWO-PARAMETER EMPIRICAL DISTRIBUTION MODELS 7-DAY AVERAGE EFFLUENT BOD$_5$

	One-Parameter Model				Two-Parameter Model				
Percentile	A_1	C_1	r^2	Standard Error	A_2	B_2	C_2	r^2	Standard Error
0	0.55	-2.8	0.75	5.3	0.83	-1.09	-1.1	0.93	3.0
10	0.62	0.3	0.83	4.7	0.90	-1.08	1.9	0.99	1.3
20	0.67	0.8	0.86	4.6	0.94	-1.06	2.4	0.99	1.3
30	0.78	0.1	0.93	3.5	1.00	-0.81	1.3	0.99	1.0
40	0.91	-1.0	0.98	2.0	1.03	-0.46	-0.3	0.99	0.7
50	1.07	-2.7	0.99	1.5	1.09	-0.05	-2.6	0.99	1.6
60	1.19	-3.1	0.99	2.0	1.12	0.27	-3.5	0.99	1.7
70	1.24	-2.1	0.98	3.1	1.11	0.53	-2.9	0.99	2.2
80	1.27	0.0	0.97	3.9	1.05	0.85	-1.3	0.99	1.7
90	1.33	2.9	0.93	6.0	0.97	1.39	0.8	0.99	1.7
95	1.32	6.0	0.91	7.1	0.89	1.64	3.5	0.99	2.0
98	1.32	9.9	0.85	9.4	0.77	2.11	6.7	0.98	3.4
99	1.31	12.5	0.76	12.2	0.62	2.67	8.5	0.96	5.4
100	1.26	17.0	0.64	15.6	0.44	3.16	12.3	0.90	8.8

EPA-287 (Cin.)
(4-76)

TABLE A-3

COEFFICIENTS AND STATISTICAL RESULTS FOR ONE- AND TWO-PARAMETER EMPIRICAL DISTRIBUTION MODELS 7-DAY AVERAGE EFFLUENT SUSPENDED SOLIDS

	One-Parameter Model				Two-Parameter Model				
Percentile	A_1	C_1	r^2	Standard Error	A_2	B_2	C_2	r^2	Standard Error
0	0.54	-3.7	.76	4.6	1.01	-1.88	0.1	0.97	1.7
10	0.73	-2.7	.96	2.1	0.92	-0.78	-1.1	0.99	1.2
20	0.77	-1.3	.97	2.0	0.96	-0.80	0.3	0.99	0.7
30	0.81	-0.3	.98	1.5	0.96	-0.60	0.9	0.99	0.7
40	0.90	-0.5	.99	1.6	1.05	-0.62	0.8	0.99	0.8
50	0.98	-0.6	.99	1.8	1.13	-0.59	0.6	0.99	1.2
60	1.05	-0.2	.99	1.2	1.12	-0.26	0.3	0.99	1.1
70	1.12	0.9	.99	1.3	1.03	0.36	0.1	0.99	1.0
80	1.18	2.2	.98	2.2	0.97	0.84	0.4	0.99	1.1
90	1.40	1.1	.96	4.6	0.93	1.89	-2.7	0.99	1.5
95	1.53	1.3	.93	6.2	0.91	2.52	-3.8	0.99	2.2
98	1.52	7.2	.90	7.6	0.85	2.73	1.7	0.97	4.6
99	1.50	10.2	.90	7.7	0.91	2.36	5.5	0.95	5.8
100	1.50	13.4	.87	8.6	0.76	2.97	7.3	0.95	5.5

EPA-287 (Cin.)
(4-76)

TABLE A-4

COEFFICIENTS AND STATISTICAL RESULTS FOR ONE- AND TWO-PARAMETER EMPIRICAL DISTRIBUTION MODELS 30-DAY AVERAGE EFFLUENT BOD_5

	One-Parameter Model				Two-Parameter Model				
Percentile	A_1	C_1	r^2	Standard Error	A_2	B_2	C_2	r^2	Standard Error
0	0.59	0.4			0.89	-1.11	0.4		
10	0.62	1.3	0.82	4.9	0.92	-1.12	1.3	0.98	1.8
20	0.68	1.6	0.90	3.9	0.94	-0.95	1.6	0.99	0.7
30	0.78	1.1	0.91	4.2	1.01	-0.88	1.1	0.97	2.4
40	0.89	0.0	0.97	2.6	1.04	-0.57	0.0	0.99	1.2
50	0.99	-1.0	0.99	1.9	1.07	-0.30	-1.0	0.99	1.4
60	1.22	-4.4	0.98	2.8	1.14	0.31	-4.3	0.99	2.6
70	1.27	-3.2	0.97	3.8	1.07	0.77	-3.2	0.99	2.3
80	1.28	-1.1	0.96	4.4	1.00	1.04	-1.1	0.99	1.4
90	1.32	2.0	0.93	6.2	0.94	1.45	2.0	0.99	2.1
95	1.33	3.9	0.92	6.8	0.90	1.61	3.9	0.99	2.1
98	1.33	5.1	0.89	7.7	0.86	1.78	5.2	0.99	2.7
99	1.31	6.4	0.88	8.1	0.81	1.86	6.5	0.99	2.9
100	1.31	7.1	0.88	8.3	0.79	1.94	7.2	0.99	2.8

TYPING GUIDE SHEET
BEGIN FIRST LINE OF TEXT HERE
CENTER OF PAGE
TOP OF IMAGE AREA
DROPPED HEAD; BEGIN SECTIONS HERE
BEGIN LAST LINE OF TEXT
BOTTOM OF IMAGE AREA OUTSIDE DIMENSION FOR TABLES AND ILLUSTRATIONS
3/8"
PAGE NUMBER
EPA-287 (Cin.)
(4-78)

TYPING GUIDE SHEET
BEGIN FIRST LINE OF TEXT HERE
CENTER OF PAGE
TOP OF IMAGE AREA
DROPPED HEAD BEGIN SECTIONS HERE

TABLE A-5

COEFFICIENTS AND STATISTICAL RESULTS FOR ONE- AND TWO-PARAMETER EMPIRICAL DISTRIBUTION MODELS 30-DAY AVERAGE EFFLUENT SUSPENDED SOLIDS

	One-Parameter Model				Two-Parameter Model				
Percentile	A_1	C_1	r^2	Standard Error	A_2	B_2	C_2	r^2	Standard Error
0	0.72	-3.9	0.95	2.6	1.00	-1.20	-3.3	0.99	0.9
10	0.77	-2.1	0.98	1.7	0.94	-0.71	-1.8	0.99	0.9
20	0.76	0.0	0.99	1.3	0.89	-0.54	0.3	0.99	0.7
30	0.77	1.6	0.98	1.5	0.91	-0.60	1.9	0.99	0.9
40	0.85	1.1	0.98	1.9	1.06	-0.91	1.5	0.99	0.6
50	0.92	1.5	0.99	1.4	1.06	-0.61	1.8	0.99	0.8
60	1.08	-0.2	0.99	1.8	1.16	-0.36	0.0	0.99	1.7
70	1.13	-0.2	0.99	1.9	1.16	-0.12	-0.1	0.99	2.0
80	1.27	-0.8	0.98	2.4	1.05	0.94	-1.3	0.99	1.5
90	1.35	0.1	0.95	4.5	0.88	2.04	-0.9	0.99	1.7
95	1.37	1.4	0.94	5.3	0.81	2.42	0.2	0.99	2.0
98	1.41	2.3	0.92	6.0	0.77	2.74	0.9	0.99	2.4
99	1.43	2.4	0.92	6.2	0.78	2.81	1.0	0.99	2.6
100	1.41	3.6	0.91	6.5	0.75	2.89	2.2	0.98	2.9

BEGIN LAST LINE OF TEXT
BOTTOM OF IMAGE AREA OUTSIDE DIMENSION FOR TABLES AND ILLUSTRATIONS
3/8"

PAGE NUMBER
EPA-287 (Cin.)
(4-76)

APPENDIX B

DEVIATIONS OF MONTHLY AVERAGES FROM ANNUAL MEAN

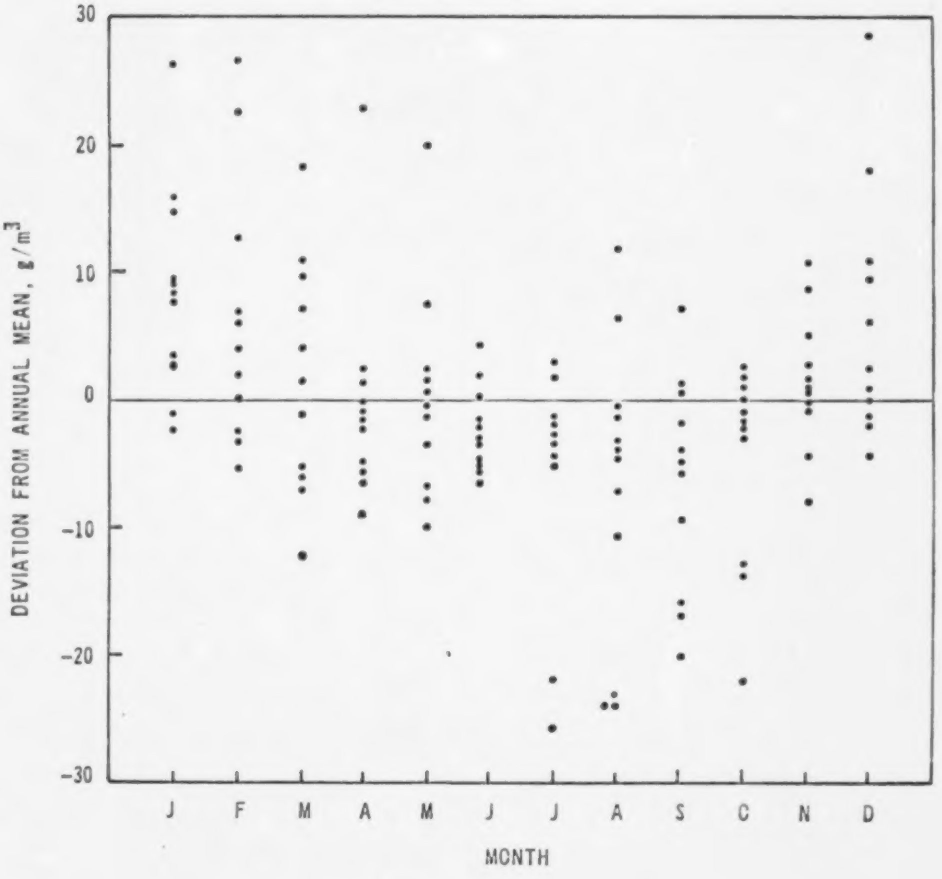


Figure B-1. **Deviations of monthly average BOD_5 concentrations from annual mean.**